Project/Work Title

The development of quantitative risk-based strategies for bovine Tuberculosis in England and Wales.

Authors and Affiliation

Prof. Rowland R. Kao (Project lead, mathematical modelling and overall report responsibility),
Dr. Liliana Salvador (Postdoctoral Research Assistant, simulation modeling, data analysis),
Mr. Michael Deason (Postdoctoral Research Assistant, data analysis),
Mr. Thomas Doherty (Research Associate, simulation modeling).

Boyd Orr Centre for Population and Ecosystem Health, Institute for Biodiversity, Animal Health and Comparative Medicine, University of Glasgow (RRK, LS, MD, TD).

Additional contributions were made by:
Dr. Jessica Enright (Postdoctoral Research Fellow, machine learning algorithms, funded under the Epic project)
Dr. Paul Bessell (Postdoctoral Research Fellow, Roslin Institute, University of Edinburgh, expert advice on the implementation of the risk-based surveillance model, funded under the Epic project)
Dr. Anthony O'Hare (Lecturer, University of Stirling, simulation modeling)
Dr. Richard Orton (Postdoctoral Research Associate, analyses of transitional areas, funded under BBSRC, project number BB/I013784/1).

Contact details of the Main Author

Prof. Rowland R. Kao
Boyd Orr Centre for Population and Ecosystem Health,
Institute of Biodiversity, Animal Health and Comparative Medicine
University of Glasgow
464 Bearsden Rd. Glasgow G61 1QH
Telephone: 0141-330-5770
Email: rowland.kao@glasgow.ac.uk

Contact Details of the Defra/APHA QM Coordinator

Dr. Susan Campbell
Animal & Plant Health Agency (APHA)
Veterinary Evidence Unit - bTB Programme
Area 4A, Nobel House
17 Smith Square
London
SW1P 3JR
Telephone: 020 7238 5513
Email: Susan.L.Campbell@ahvla.gsi.gov.uk

Executive Summary (Overall Findings/Results)

The aim of this project is to provide recommendations for improved surveillance for bovine Tuberculosis (bTB) in English and Welsh cattle, considering three important, distinct problems:

1. In areas where there is a low incidence (Low Risk Areas or LRAs) of bTB in cattle and the default testing regime is currently quadrennial, to identify approaches that will reduce the total testing burden, without compromising the detection of herds containing infected cattle and efficiently identify high risk herds, possibily for additional testing.

2. In areas where there is a high incidence (High Risk Areas or HRAs) of bTB in cattle, to identify approaches that more quickly identify herds which may present a higher risk of onward transmission to other herds.

3. In areas which are currently perceived as LRAs but are in danger of becoming HRAs (Transition Areas or TAs), to identify signatures of this transition, with an aim of informing increasing surveillance (i.e. moving from (1) above to (2)) and or systematically curtailing spread of HRAs themselves.

The work to achieve these objectives was delivered in four separate stages:

1. **Identification of areas that could be epidemiologically classified as HRA, LRA and TA**. Conservative approaches were used to identify HRAs and LRAs, as these were considered to be restricted to at most only those that were consistently under annual testing over the timeframe to be HRAs, and for LRAs, considering only areas currently under quadrennial testing (i.e. since the boundaries of quadrennial testing areas were simplified in 2013). The identification of TAs is more complex, as the changes in testing regime over time do not necessarily reflect changing epidemiological risk in a quantifiable fashion, and blur the epidemiological signal of increasing outbreaks. Therefore, using a bespoke statistical model that balances off the two factors of 'local risk' and movement-based risk, for each year from 2003 to 2007, areas were identified (down to a 10 km $\times$ 10 km grid square or 'quadrat' resolution) that were statistically most likely for local risk to be important, considering the spread of specific, genetically closely related groups (by 'spoligotype') of outbreaks. Quadrats that were consistently identified as having identified local risk over the entire timeframe were considered HRAs, those which never were identified as having a local risk were considered LRAs, and those which were intermittently considered to have a local risk becoming infected in the last year of the study were considered TAs.

2. **Identification of statistical risk factors**. In each area, the risk factors for breakdowns were identified using a combination of statistical models. The following key outcomes are highlighted:

   (a) Finishing herds increase the risk in HRAs, however, in LRAs they appear to have a similar risk profile to beef. This suggests that increased breakdowns in finishing herds may be a good indicator of an incipient local problem.

   (b) Movements of cattle from HRAs are significant risk factors for the recent years of LRAs, HRAs and for TAs. While they increase the risk in low risk areas, the results suggest that they do not contribute to the increase of risk in HRAs and TAs.

   (c) Imports from Ireland became a significant risk factor in the recent years in LRAs, but no evidence of being important in other areas.

(d) While risk factors suggest that large herds are important in all areas, it is in LRAs where they have increased risk. Large herds will, in principle, have a more intensive inter-herd dynamic by selling and buying more animals, increasing their risk of receiving infected animals from HRAs and Ireland.

(e) There is statistical evidence that imputed badger density (and an interaction between imputed badger density and beef herds) might be a significant risk factor in LRAs for the period 2008-2013, however, given the known epidemiology of the disease, it is likely that this is a proxy for other, geographically related correlates. This also suggests that the imputed badger densities used here, need to be interpreted with caution.

The aim was to identify what signatures distinguished the three types of areas.

3. **Identifying outcomes of revised surveillance strategies.** With a previously developed surveillance model that was used to identify targeted strategies in Scotland, for herds in LRAs in England, a variety of testing regimes were considered with the aims of identifying the impact on total breakdowns under better, similar or reduced surveillance levels (for full surveillance results, see tables 12 to 17 in the main report). Our analyses suggest that:

(a) Slaughterhouse surveillance alone would result in substantially fewer detected infections than the current system (140 compared to 161 in 2003-2008; 117 compared to 145 in 2009-2013), which could increase the potential for onward transmission in LRAs and enhance the risk of establishment of new HRAs.

(b) There appears to be little or no advantage to implementing additional testing in LRA England (i.e. above the *four-year* testing scenario) since the probability of picking up additional infections is very low. The *one-year* and *two-year* baseline scenarios show that the number of detected infections is very similar to the *four-year* regime (154 and 159 compared to 161 in 2003-2008; 137 and 143 compared to 145 in 2009-2013), and they do present fewer latent infections (5 and 15 compared to 31 in 2003-2008; 4 and 14 compared to 31 in 2009-2013), but require additional herd tests (19564 and 9782 compared to 4891 in 2003-2008; 12939 and 6469 compared to 3234 in 2009-2013) and result in more false positives (433 and 216 compared to 108 in 2003-2008; 287 and 143 compared to 71 in 2009-2013).

(c) A number of the tested scenarios offer clear advantages over the others. In *scenarios 1* and *2*, some herds are exempt from routine herd testing (RHT) due to sending many cattle to slaughter, but others are tested more frequently due to movements of cattle from high risk areas in England and Wales, and due to Irish imports. *Scenarios 1-2*, when compared to the current regime, present similar or better levels of detection of breakdowns (163 and 160 compared to 160 in 2003-2008; 148 and 145 compared to 145 in 2009-2013), a similar number of mean latent infections (30 and 33 compared to 31 in 2003-2008; 28 and 32 compared to 31 in 2009-2013), and present fewer false positives (94 and 81 compared to 108 in 2003-2008; 68 and 55 compared to 71 in 2009-2013). They test fewer herds (4262 and 3729 compared to 4891 in 2003-2008; 3064 and 2531 compared to 3234 in 2009-2013) than the current regime, and they can detect on average 82% of the breakdowns identified by RHT (16 out of 21 in 2003-2008; 8 out of 9 in 2009-2013). In contrast, while *scenarios 7-8* (which include more frequent RHT of high risk herds, and no exemptions) only offer minimal benefits for the additional tests done. There are very similar levels of detection of breakdowns (162 and 161 compared to 161 in 2003-2008; 146 and 145 compared to 145 in 2009-2013) and a similar number of latent infections (30 in both scenarios compared to 31 in both 2003-2008 and 2009-2013) when compared to the current regime. However, they do present a higher number of false positives (125 and 118 compared to 108 in 2003-2008; 84 and 80 compared to 71 in 2009-2013) and a higher number of herds tested (5427 and 5173 compared to 4891 in 2003-2008; 3626 and 3486 compared to 3234 in 2009-2013) than the current regime. Because there is no

exemption from testing in *scenarios 7-8*, these can detect 100% of the breakdowns identified under current testing regimes.

(d) In LRA England, *scenarios 3-4* that penalize herds that slaughter small percentages of their stock and have been importing animals from risky areas for a long period (and have more than 100 animals - *scenario 4* only) offer a saving of the number of herds tested during both time periods when compared to the current scenario (2823 and 3301 compared to 4891 in 2003-2008; 2074 and 2827 compared to 3234 in 2009-2013), and fewer false positives (61 and 81 compared to 108 in 2003-2008; 45 and 68 compared to 71 in 2009-2013), however, they detect slightly fewer infected herds (158 and 156 compared to 161 in 2003-2008; 143 in both scenarios compared to 145 in 2009-2013), and show a greater number of latent infections (36 and 38 compared to 31 in 2003-2008; 34 and 33 compared to 31 in 2009-2013) when compared to the current scenario. *Scenarios 5-6* are a replicate of *scenario 4* but with cutoffs for testing exemptions occurring only for comparatively larger herds. These scenarios offer a saving of the number of herds tested during both periods when compared to the current scenario (1811 and 1654 compared to 4891 in 2003-2008; 1803 and 1682 compared to 3234 in 2009-2013) and produce fewer false positives (41 and 35 compared to 108 in 2003-2008; 41 and 37 compared to 71 in 2009-2013), however, they detect fewer infected herds (151 and 150 compared to 161 in 2003-2008; 140 for both scenarios compared to 144 in 2009-2013), and they show a greater number of latent infections (44 and 45 compared to 31 in 2003-2008; 38 and 39 compared to 31 in 2009-2013).

4. Within herd models of bTB transmission identified substantial detection delays that might result in breakdowns should infections be seeded in herds exempt from RHT. While this is not likely to create onward consequences should RHT exempt herds of sending cattle to other herds, the possibility of onward infection otherwise needs to be considered carefully, possibly requiring additional measures such as enhanced pre-movement testing. These models showed that dairy and beef herds in HRAs show similar epidemiological characteristics with no evidence that one or the other presents substantially more inherent individual level risk of cattle-to-cattle transmission (i.e. beyond risks due to high levels trading), though they suggest that dairy herds may experience a higher level of risk due to external infection sources.

   While we view these recommendations as robust given the epidemiological knowledge and data available to us, a critical consideration when reviewing these recommendations is the knowledge that the landscape of bTB transmission to cattle herds in GB is always changing, because of natural factors (the epidemiology of bTB in cattle and badgers), changes in the industry, and due to responses to changes in legislation. As such, while in our view the principles behind the recommendations are sound, the specific quantitative aspects of the recommendations would require regular review and re-analysis in order to remain current to any ongoing bTB epidemic.

Introduction

Background

Bovine Tuberculosis (bTB) in cattle is the most economically important and controversial disease of livestock in Great Britain, with important impacts on animal health and welfare, and farmers' livelihoods and well-being. The primary requirement of surveillance for bovine Tuberculosis in British cattle is the timely identification of infected cattle in order to reduce onward infection (particularly between herds) with the aim to contain, control and if possible, ultimately reduce the incidence due to cattle-to-cattle transmission. Current testing regimes for bTB, while already sensitive to some differences in herd risk status, could be enhanced by a more targeted approach that considers the epidemiological evidence of risk found in the comprehensive data on bTB in British cattle, in order to either reduce testing of cattle (thereby saving cost and effort), or enhancing identification of infection (thereby reducing onward risk), or both.

There are three primary means by which targeted surveillance might improve the testing/detection balance in Great Britain:

1. In areas where cattle are clearly at a low risk of infection with bTB, a reduction of testing, if it does not compromise detection, will both reduce costs to government and reduce the burden on farmers. This is the approach used in Scotland, where only herds at a sufficiently high infection risk, mostly due to imports of cattle from areas at high risk of bTB, are tested regularly. Some improvements may also be achieved by more intensive testing of higher risk herds.

2. In areas where there is a high risk of infection with bTB, enhanced testing of herds that are more likely to spread bTB to other cattle could detect these more rapidly, and thereby, reduce onward spread.

3. In areas where there is believed to be a low risk of infection, but there is danger of transition due to the spread of high risk areas. Improvement on surveillance to identify this transition will both identify where surveillance needs to be enhanced by further testing, and potentially prevent high risk areas from spreading.

The existing bTB surveillance regime already includes some elements of risk-based targeting of surveillance, for example in the form of different minimum frequencies of routine herd testing (every 1 or 4 years) with those under 4 year testing largely assured of being low risk. Additionally, certain herd types that are perceived to be at higher risk of acquiring infection are subject to the most stringent testing frequency (annual) regardless of their location. Pre-movement bTB testing also targets herds situated in the endemic/high risk bTB regions. This project aims to identify further means of improving targeted surveillance, considering all of the above three circumstances.

The expansion of areas in Great Britain (GB) with a high incidence of herd breakdowns of bTB remains largely unexplained and poorly quantified. The crux of this proposal is a new and epidemiologically robust way of defining the manner in which areas are defined as high or low risk, with areas that become high risk over a specified time frame, identified as transitional. Based on historical data, and modifying an approach originally developed in a previous Defra project, we have identified areas which were transitional in the analysed time frame (concentrating on LRAs that have become persistent HRAs). By analysing risk factors (including imputed badger presence/absence data) associated with each of the area types (high risk, low risk and transitional), we aim to address the risk-based surveillance requirements across all three. Since breakdowns can be seeded a considerable time prior to detection,

the difference in testing regime times can exacerbate the establishment of new areas where cattle herds are at a higher risk of breakdown (HRAs).

Purpose of the work

1. Development and analysis of locations of HRAs and TAs
2. Statistical analysis of underlying risk factors for breakdowns in HRA, LRAs and TAs
3. Application of the Scottish risk-based surveillance model to long term LRAs to reduce testing
4. Use mathematical models to determine consequences of possible missed breakdowns

Methods

1. **Development and analysis of locations of HRAs and TAs**

Source data
The source data used for this analysis were extracted from the Cattle Tracing System (CTS) of GB, provided by RADAR (Rapid Analysis and Detection of Animal-related Risks), and details of bTB breakdowns were extracted from DEFRA's animal health database, VetNet. Cattle movements are represented as daily reported links between pairs of CTS herds using cattle ID numbers. Breakdowns were identified according to date and CPH from confirmed cases in the VetNet data. Genotype data consisting of spoligpotypes and VNTR types of *M. bovis* were obtained from the Animal & Plant Health Agency (APHA). Three of the provided genotypes were chosen for investigation on the basis of its geographical predominance in expanding regions of high-incidence (used in the model developed by [13] on the prediction of high-risk areas of bTB). These were genotypes 25:a (spoligotype 25 (SB0129), VNTR type $6-5-5-4^*-2-3.1$ mainly prevalent in the English Midlands around Stoke-on-Trent), genotype $10:\ a$ (spoligotype 10 (SB0272), VNTR type $7-5-5-4^*-3-3.1$ on the Welsh borders) and genotype $9:\ b$ (spoligotype 9 (SB0140), VNTR type $7-5-5-5^*-3-2.1$, in SW Wales). Each of three genotype datasets were linked to breakdowns in VetNet via cattle eartag numbers to generate a list of herd breakdowns for each genotype.

Identification and characterization of genotype-specific TAs
In the work developed by Orton and Colleagues ([13] under Defra Project SE3243), GB was divided into 10×10 km quadrats and each spoligotype was assigned to each quadrat (see Figure 1).

That analysis was refined and expanded here, with a three-phase process used to identify HRAs in terms of these quadrats:

- In the first phase, each quadrant for a specific genotype was used individually in a simulation as the sole HRA to explain cattle herd breakdowns for that genotype, considering both presence in the quadrat and movements of cattle out of it, and the likelihood of the overall spoligotype specific epidemic recorded. Any quadrant with more than two herds breaking down with the selected genotype during the study period 2003-2007 were considered for evaluation as an HRA. Epidemic likelihoods were obtained from Nelmer-Mead simplex fits, with the best likelihood observed used to rank the quadrats.
- In the second phase of model selection, these quadrats were sequentially aggregated into a larger HRA, considering the quadrant with the likelihood from phase one, in order down to the lowest likelihood. The model likelihood was reevaluated with each quadrant addition and the result recorded. A plot of the change in likelihood would be expected to show an overall initial increase in likelihood (since a single quadrat would be unlikely to be the only HRA), but then eventually start to decline as too many quadrats become included as HRAs.

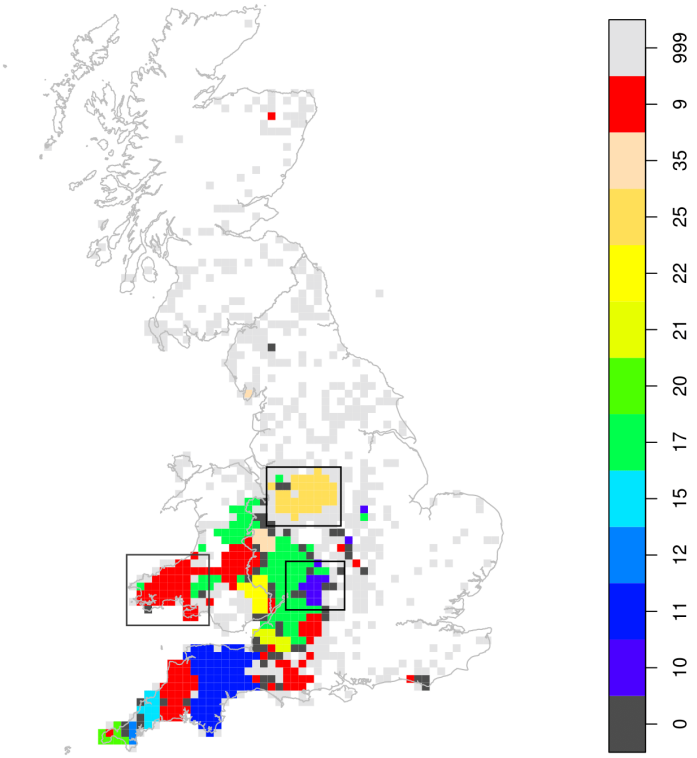


Figure 1: *M. bovis* spoligotypes in cattle by quadrat of 10 × 10 Km for GB, based on all available data to 2010. Quadrats where more than 50% of typed isolates belong to an individual spoligotype are indicated by colour, with category "999" indicating insuficient data (< 6 brekadowns) and "0" (dark grey) representing inconclusives - ie. greater than 5 breakdowns but without a dominant spoligotype. In our study, we focus on the spoligotyes highlighted with a square: S9b (red), S10a (blue) and S25a (orange).

- In the third phase of the model evaluation, in order to improve the method used in the second phase, the quadrats are reordered in order to smooth the curve. E.g. if the original order is 1-2-3-4-5, and the likelihood when quadrats 1-2 are evaluated is lower than for quadrat 1 alone, and for 1-2-3-4 lower than for 1-2-3, the quadrats are reordered as 1-3-5-2-4 and the likelihoods reevaulated as quadrats are accumulated.

This process was repeated for five different three year evaluation windows, for the years 2003-2007 (see Figure 2).

Definition of TAs. Figure 2 shows that these quadrats show considerable variability in the number of areas considered high-risk for the period 2003-2007. A quadrat was considered transitional if it was intermittently considered to have a local risk becoming infected in the last year of the study period.

2. **Statistical analysis of underlying risk factors for breakdowns in HRAs, LRAs and TAs.**

   The basis for this analysis is largely standard statistical GLM models run in the R programming environment. In addition, established machine learning approaches are used as exible methods to determine the relative predictive power of cattle movements to determine the risk of breakdowns in LRAs.

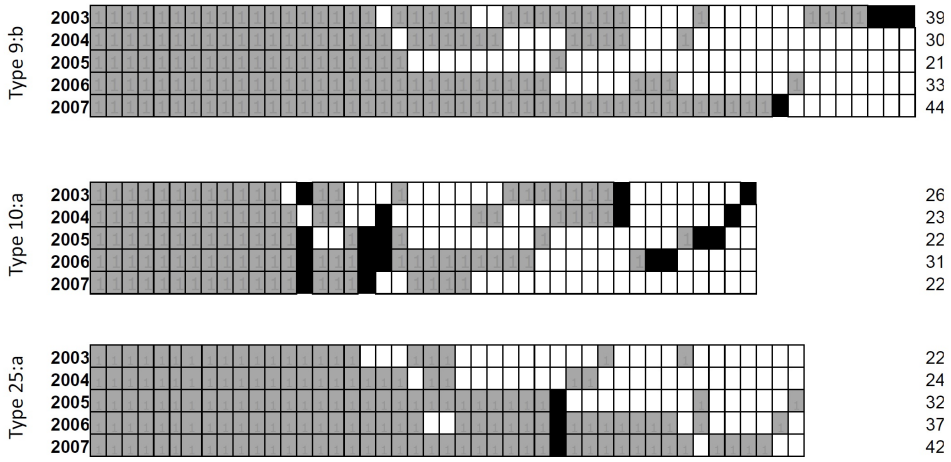


Figure 2: All quadrats for the genotypes S9b, S10a, S25a showing years in which they are designated high risk. Quadrats that appear within the largest contiguous group are in grey, those that appear in isolation (outside of the largest group) in black. Only type 25:a shows evidence of clear HRA expansion, while the HRAs for types 9:b and 10:a appears to fluctuate without covering substantially greater areas or with clear signs of steady expansion. Therefore an interesting comparison is whether type 25:a shows different underlying risk factors, compared to the other two genotype 'transitional' areas.

2.1. Risk factors as predictors of bTB breakdowns in Scotland and LRA England.

Source data

In this work we combine two datasets: the Cattle Tracing System (CTS) movement data (RADAR), and the VetNet records of bTB tests and breakdowns (DEFRA). The CTS database contains records of all inwards and outward cattle movements in Great Britain. We used a version of the dataset with markets stripped from movements: a movement from $farm_1$ to a market to $farm_2$ is recorded as a movement from $farm_1$ to $farm_2$.

We use CTS movements and VetNet breakdowns to generate a variety of risk ratings for each cattle movement from 2006 to 2011 and a risk rating for each farm based on the ratings of their inward movements for each year of the study period.

Movement risk ratings

Let $m_{ijt} = 1$ be a movement if farm i to farm j on data t. Let $N(m_{ijt})$ be the number of animals moved on that day between those farms. For each movement, four different risk ratings were calculated:

- $R_{county}(m_{ijt}) = 1$ if farm i is in 1-year testing county, 0 otherwise.
- $R_{num}(m_{ijt}) = k$ where k is the number of animals moved if i has had a bTB breakdown within ten years previous to t, 0 otherwise.
- $R_{num,time}(m_{ijt}) = k \ln(t - t_B)$ where k is the number of animals moved and t_B is the date of the most recent breakdown at i if i has had a bTB breakdown within ten years previous to t, 0 otherwise.
- $R_{time}(m_{ijt}) = \ln(t - t_B)$ where t_B is the date of the most recent breakdown at i if i has had a bTB breakdown within ten years previous to t, 0 otherwise.

Farm risk ratings
For farm j in year y:

- $R_{county}(i) = \sum R_{county}(m_{ijt})$ where t is within two years previous to y.
- $R_{num}(i) = \sum R_{num}(m_{ijt})$ where t is within two years previous to y.
- $R_{num,time}(i) = \sum R_{num,time}(m_{ijt})$ where t is within two years previous to y.
- $R_{time}(i) = \sum R_{time}(m_{ijt})$ where t is within two years previous to y.

Stump classifiers
A *stump classifier* is a simple threshold classifier. Given a threshold and a set of rated examples (in this case, farms with risk ratings), we classified all examples with ratings above the threshold as positive, and below the threshold as negative. For a given rating measure, we choose a threshold, and we classify everything below the threshold, as non-risky, and everything above as risky.

Figure 3: Example of a stump classifier. A stump classifier is a machine-learning model consisting of a one-level decision tree. A decision stump makes a prediction based on the value of a just single input feature.

At each threshold, we compare the results with bTB breakdowns and record the number of true positives and false positives. The curve joining these points is a *Receiver Operator Characteristic* (ROC) curve (Figure 4), therefore, each curve represents the analysis of one risk rating for one given year.

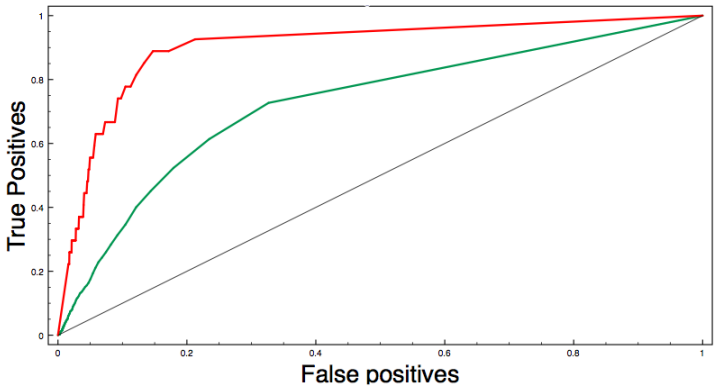


Figure 4: ROC curve. A measure of true and false positive rates over all the thresholds to assess the performance of the diagnostic test (perfect classifier: upper left hand corner; no-information random classifier: diagonal line).

These curves are commonly used in machine learning and radiology (for example) to assess the performance of a classifier or diagnostic test. A perfect classifier would be in the upper left hand corner of the plot, a no-information random classifier would be on the diagonal line from bottom left to top right corner.

2.2. Statistical analysis of underlying risk factors for breakdowns in LRAs, HRAs and TAs

Source data

In order to calculate the herd-level prevalence of infection, risk factors for herds in England and Wales with confirmed bTB test reactors were modelled. Herd attributes including herd type and location were collated into tables for the periods of interest (2002-2008 and 2008-2013) while historic bTB test results and incidence data were extracted from SAM's database provided by the Animal & Plant Health Agency (APHA). The aforementioned tables were combined in order to create the large base datasets used to construct the statistical models.

The births, deaths, imports and movements of individual cattle were extracted from the British Cattle Movement System's Cattle Tracing System (CTS).

Average holding size for each year was calculated using CTS by incorporating all recorded cattle movements off and on to individual herds. The result of these calculations was provided by APHA.

Additional data on the historical SICCT test interval data by parish was used in order to differentiate low incidence and high incidence herds through time. Only those herds that were in entirely low or in entirely high incidence areas during the time periods of interest were considered in the risk factor analyses.

Data on herds located in TAs were also used ([13]). Transitional areas of England and Wales from 2003-2007 were identified using *M. bovis* spoligotype data provided by APHA. These areas were divided into 100 km^2 quadrats and assigned incidence information, which is based on the predominant testing interval in the area and on a bespoke statistical model used to identify quadrats where a locally based risk was likely. The data set was further subdivided by the number of years of the potential five each quadrat spent in high incidence areas.

Data on the density of badgers were derived from work presented in [14]. The authors estimated the density of main badger setts for all the cells (resolution 100×100 m) in England and Wales by using the *Mahalanobis* distance statistic (D2), which is a presence only modeling technique. To do so, they have used stratified random samples of sett locations in combination with biologically relevant landscape data. The imputed badger density values were added to the herd dataset by matching for the location of each herd.

Statistical analysis

Multivariate logistic mixed models to assess candidate risk factor were formulated in which the outcome was 1 if herd i in county j recorded one or more confirmed incidences of bTB in year k and 0 if there was not. Models were constructed using data from 2002-2008 and 2008-2013 for both low and high risk areas as well as from 2003-2007 for the transitional areas. Both year and county acted as random effects in order to estimate the mean distribution of the outcome in every permutation of year and county as well as to control for any influence these values may have on estimating herd level prevalence of infection.

The following fixed effects were considered for the analyses:

(a) The mean number of cattle on a holding for each year. This variable was further categorised as 1-50, 51-100, 101-200, 201-350, 351-500 and >500 animals, corresponding with those categories used in the RADAR 2008 Cattle Book.

(b) The consolidated herd types found in SAM's database. These types were reduced to beef, dairy, finishing, suckler and stores according to Table 1. An additional group including herds that did not fall into the aforementioned categories (Other category) was excluded due to the poor representation of these herds in the bTB incidence database.

(c) A binary response indicating whether a holding imported any cattle from either the Republic of Ireland or Northern Ireland in the previous year.

(d) The number of batches of cattle arriving from high incidence areas of England and Wales in the previous year. This value was categorised as 0, 1-10, and >10.

(e) The predicted imputed density of badgers per km^2. This value was centred and divided by two standard deviations to normalise its distribution in order to facilitate model fitting.

General Herd Type	Specific Herd Categories
Beef	Beef, BEEF, Beef Bull Hirer, Beef Dealer, Beef Heifer Rearer, BISON, Bull Beef, Dealer, DEALER, Meat
Dairy	Dairy, DAIRY, Dairy Bull Hirer, Dairy Dealer, Dairy Heifer Rearer, DAIRY P/R, DAIRY/DEAL, Domestic, Producer, Producer/Caterer, Producer/Processor, Producer/Retailer, Producer/Wholesaler
Stores	Stores
Finishing	Finishing
Suckler	Suckler, BEEF SUCKLER
Other	All the other categories in SAM's DB

Table 1: Characterization of herd types in their sub-categories according to SAM's database. The 'Other' category is not used in our analyses.

Models were evaluated using the AIC model selection score, the most appropriate information criteria in models where false negatives would be more misleading than false positives. The best fit models were chosen by back-fitting the listed fixed effects, forward-fitting the random effects (county and year) and finally back-fitting the fixed effects once more to recheck the selected variables. This rechecking is done because the inclusion of random effects may, sometimes, render certain fixed effects no longer significant. The significant fixed effects are presented as odds ratios (Odds) with 95% confidence intervals. The data were prepared using a combination of UNIX shell commands for basic text editing and R for data management. The multivariate logistic mixed models were constructed in the lme4 package in R ([17]). The best fit models were chosen using fitLMER.fnc, which is part of the R package LMERConvenienceFunctions ([18]).

3. **Application of the Scottish risk-based surveillance model to long term LRAs to reduce testing**

The aims of these analyses are the following:

- Evaluate whether slaughterhouse surveillance alone is sufficient to ensure a high level of confidence of freedom from infection
- Evaluate alternative strategies to RHT that are based upon risk. This should incorporate both the risk of infection and the risk of detection at slaughterhouse
- Minimize the number of false-positive herds identified by testing

These analyses used the mixed logistic regression model from section 2 to define the probability of a herd becoming infected, and implemented components of a stochastic simulation model developed by APHA (then AHVLA, the meta-analysis study team, personal communication), to calculate the probability of a herd's freedom from infection. Data from the original Scottish study (covering the period 2003-2008) was compared to data from England and Wales over the same period, and to data from England and Wales to a more recent period (2009-2013).

Source data
The data used for this analysis were derived from SAM's and CTS. The following steps were used to derive the cattle herd data:

(a) All herds with a unique CPH number on the SAM's herd table that were active during all of the years of each study period (2002-2008 and 2008-2013) and located in LRAs were identified.

(b) Of the herds identified above, only those that had animals recorded on CTS were included. For these, the average number of animals in the herd per year was calculated. Only the ones that had herd size > 0 were included.

(c) Herds with a unique CPH but with multiple herd types were excluded from the analysis.

(d) One spatial location was chosen at random for herds with a unique CPH that had multiple database entries with different spatial locations.

(e) The number of animals sent to slaughter from these herds in each year was calculated.

During 2003-2008, all eligible herds in Scotland were tested once every 4 years under the routine herd testing (RHT) policy and supplemented by further herd tests: *post-movement* and *post-Irish import* triggered by moving, respectively, cattle from HRA England and Wales, and Ireland into Scotland, contact *tracing* following a confirmed breakdown, and retest of animals that had inconclusive results during their SICCT test (60 days later) (Table 9).

During 2003-2008 and 2009-2013, all eligible herds in LRA England were also tested once every 4 years under the routine herd testing (RHT) policy and supplemented by the following tests (not an exhaustive list):

- *Contiguous herd* carried out on herds contiguous to OTFW (Officially Tuberculosis Free status withdrawn) herds.
- *1st hotspot check* carried out if a holding is within a 3 km radius of an incident, which triggers a potential hotspot area.
- *Pre-movement* test to be carried out 60 days or less prior to movement of animals from an annually tested herd.
- *Whole-Herd-Testing (WHT)*, which are carried out routinely every 12 months in annual testing areas and in individual herds requiring annual testing, e.g. producer-retailer dairy herds, bull hirers, heifer rearers, city/open farms, AI centres, etc. Can also be carried out via Regional Veterinary Lead (RVL) discretion in 48 month testing areas.

- *Tracing* carried out outside the normal testing frequency for the herd, to determine its disease status when there is a suspicion of infection. This will follow the back tracing of OTFW reactors found in another herd.
- *Post-movement* test to be carried out in circumstances where cattle have been moved into a holding without a required pre-movement test.
- *Post-Irish* import test carried on cattle imported from Northern Ireland and the Republic of Ireland 60 days after arrival.
- *Inconclusive reactor retest* carried out on inconclusive reactors identified at an earlier test. To be conducted at least 60 days later.

Model description

To evaluate the likelihood of herd-level freedom from infection with bTB during a specified time period (t) the model requires that the following parameters are defined:

(a) The probability of the herd becoming infected during t ($p(intro)$). This is derived from the regression model in the previous section.

(b) The number of animals in the herd (N).

(c) The bTB surveillance implemented on the farm:

  i. Slaughterhouse

  ii. Whole herd testing (i.e. testing the entire herd)

  iii. The herd-level prevalence of infection p_{star}

The efficacy of the surveillance system is evaluated by calculating the herd-level test system sensitivity se_{system}, which includes the routine whole herd SICCT testing and part herd slaughterhouse testing:

$$se_{system} = 1 - (1 - se_{herd})(1 - se_{part}),$$

in which se_{herd} is the sensitivity of the SICCT implemented as a herd test, and se_{part} is the part herd sensitivity for slaughterhouse surveillance. The herd sensitivity for a whole herd test is calculated as:

$$se_{herd} = 1 - (1 - se_{SICCT})^d,$$

in which d is the number of infected animals in the herd defined as:

$$d = N \times p_{star}.$$

The value d is derived from the product of a beta(2,90) distribution and the annual average number of animals in the herd provided by APHA. The sensitivity for a part herd test for the proportion of the herd that is sent to slaughterhouse is:

$$se_{part} = 1 - (1 - \frac{n \times se_{slh}}{N})^d,$$

where n is the number of animals sent to slaughter. In this analysis, the distributions of the test sensitivities were defined by a meta-analysis performed by the APHA meta-analysis study team (personal communication; [21]: $se_{SICCT} = 69.30\%$ and $se_{herd} = 51.11\%$. The specificity of the whole herd test is defined as follows:

$$sp_{herd} = 1 - sp_{animal}^n,$$

where $n = N$ for whole herd tests and sp_{animal} is the specificity of the test (99.58%). The probability of freedom (the posterior) at t is given by:

$$p(free) = \frac{1 - prior_t}{(1 - prior_t) + prior_t \times (1 - se_{system})},$$

where $prior_t$ is the prior probability that the herd is infected. The prior for $t+1$ is:

$$prior_{t+1} = ((1 - p(free)_t) + p(intro)) - ((1 - p(free)_t) \times p(intro)).$$

The model was implemented in the R statistical environment and run for 100 simulations. The model was implemented for all eligible herds in LRA England for the years between 2002-2008 and 2008-2013. The extra year at the beginning of the study period was to enable a 'burn-in' period for the model to ensure that it was stable for the period of the simulation. For whole herd tests over a regular repeated period (such as four-year testing) the start year of the herd testing cycle was generated randomly for each iteration.

The following were used as likely determinants of the risk of infection and subsequent detection:

(a) The size of herds - larger herds being at greater risk of infection.

(b) The proportion of the herd's total stock that is sent to slaughter during each time period - herds that send less stock to the slaughterhouse require more surveillance.

(c) Where the holding sources its stock - whether the holding is buying in animals from HRA in England, Wales, Republic of Ireland and Northern Ireland.

Four different baseline scenarios can be modelled based upon an annual time-frame for surveillance and assuming that slaughterhouse surveillance will continue:

(a) Slaughterhouse surveillance only

(b) Four year testing: WHT for all herds for every 4 years and slaughterhouse surveillance

(c) Two year testing: WHT for all herds for every 2 years and slaughterhouse surveillance

(d) One year testing: annual WHT for all herds and slaughterhouse surveillance

The following 8 risk-based scenarios below were also simulated. Here, the surveillance system is based on a point scoring system, with points added or subtracted based on perceived increases or reductions in risk.

- Scenario 1:
  - Slaughtering 25% of stock (+1)
  - Receiving 'high-risk' animals in > 3 years and slaughtering $< 50\%$ of stock (+1)
- Scenario 2:
  - Slaughtering $< 25\%$ of stock and/or receiving 'high risk' animals in > 3 years and slaughtering $< 40\%$ of stock (1)
- Scenario 3:
  - Slaughtering $< 12.5\%$ of stock and/or receiving 'high risk' animals in > 3 years and slaughtering $< 25\%$ of stock (1)
- Scenario 4:
  - Slaughtering $> 25\%$ of stock (-1)
  - Slaughtering $< 5\%$ of stock (+1)
  - Receiving 'high risk' animals in > 3 years (+1)
  - Having > 100 animals (+1)
- Scenario 5:
  - Slaughtering $> 25\%$ of stock (-1)
  - Slaughtering $< 5\%$ of stock (+1)
  - Receiving 'high risk' animals in > 3years (+1)

  - Having $>$ 350 animals (+1)
- Scenario 6:
  - Slaughtering $>$ 25% of stock (-1)
  - Slaughtering $<$ 5% of stock (+1)
  - Receiving 'high risk' animals in $>$ 3years (+1)
  - Having $>$ 500 animals (+1)
- Scenario 7:
  - Having $>$ 350 animals (3)
  - Having $\leq$ 350 animals (1)
- Scenario 8:
  - Having $>$ 500 animals (3)
  - Having $\leq$ 500 animals (1)

Model evaluation
The risk-based scenarios were evaluated for the time frames 2003-2008 and 2009-2013 by comparing the following variables to the equivalent fitted values from the *four-year* RHT surveillance:

(a) The annual number of herds tested

(b) The annual number of cattle tested

(c) The annual fitted number of detected infections ($prior_t - p(free)_t$)

(d) The annual number of latent infections ($1 - p(free)_t$)

(e) The annual expected number of false positives (sp_{herd})

A false positive is defined as an unconfirmed reactor.

4. **Use mathematical models to determine consequences of possible missed breakdowns**

In order to evaluate evidence for differences in within-herd transmission characteristics across different herd types, the within-herd transmission of bTB is modelled as a hierarchy of nested model structures, where, in the simplest case, cattle are either *susceptible*, *exposed*, *test sensitive*, or *infectious*, and models fitted to outbreak data for (in the first instance) dairy vs. beef cattle breakdowns. As well as infection structure, risk group specific age structure was accounted for, with each infectious stage further split into N_A separate age groups.

Susceptible cattle become *exposed* through infectious contact within the herd, and through external factors that may include for example, inward cattle movements, contiguous spread from neighbouring herds, or the presence of a wildlife reservoir. These external factors are incorporated into the model via a single force of infection. The model is depicted schematically in Figure 5. Once an animal becomes infectious it remains so until it is detected, at which point the animal would be culled.

We allow for heterogeneity in the infectiousness of individuals by incorporating 'super-spreaders'; i.e. with some individuals more likely than average to infect others if, for example, it excretes more bacteria than average, as suggested by experimental data [1]. A fraction of the herd, P_S, are modelled as super-spreaders where the transmission term is scaled by a factor ζ_S.

The average infectiousness of individuals may vary between herds [2] and this is incorporated into the model by allowing transmission of the disease for all livestock to be scaled by a factor ζ_H in a proportion of herds, P_H, and similarly we consider two levels of variability in α.

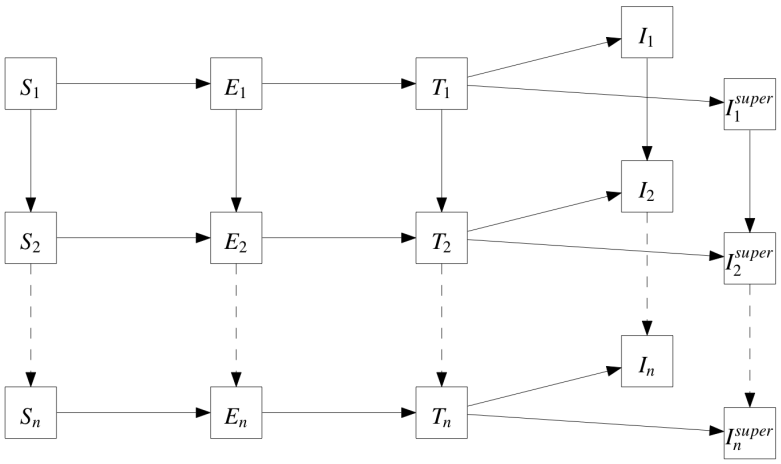


Figure 5: Disease propagation though the infection stages, *susceptible*, *exposed*, *test sensitive* and *infectious* in the age-based SETI model. Individuals move either into a new infection stage (horizontally) or to the next age bracket (vertically) as denoted by subscripts. There are two classes of infectious individuals; super-spreaders (when included in the model) are categorised as super-infectious where the transmission parameter is scaled by the value ζ.

The deterministic model is written as an ODE

$$\begin{aligned}
\frac{dS_i}{dt} &= -\beta S_i I - \zeta_S \beta S_i I^{super} - \nu_i S_i + \aleph_i \left(S_{i-1} - S_i\right)|_{i>1} - \alpha S_i \\
\frac{dE_i}{dt} &= \beta S_i I + \zeta_S \beta S_i I^{super} - \sigma E_i - \nu_i E_i + \aleph_i \left(E_{i-1} - E_i\right)|_{i>1} + \alpha S_i \\
\frac{dT_i}{dt} &= \sigma E_i - \gamma T_i - \nu_i T_i + \aleph_i \left(T_{i-1} - T_i\right)|_{i>1} \\
\frac{dI_i}{dt} &= (1 - P_S)\,\gamma T_i - \nu_i I_i + \aleph_i \left(I_{i-1} - I_i\right)|_{i>1} \\
\frac{dI_i^{super}}{dt} &= P_S \gamma T_i - \nu_i I_i^{super} + \aleph_i \left(I_{i-1}^{super} - I_i^{super}\right)|_{i>1}
\end{aligned} \tag{1}$$

where subscripts denote the age group, infection states with no subscript means the sum over all groups e.g. $I = \sum_i I_i$ and α is the force of infection external to the herd. The parameters $\aleph_i$ is the rate cattle in one age group move up to the next (older) age group so that $\aleph_i S_{i-1}(t)$ is the number of cattle entering the $S_i(t)$ state and $\aleph_i S_i(t)$ the number leaving. Cattle in the oldest age group are removed from the system at the rate $\aleph_{N_A}$. Cattle are also removed from the herd (via death or export) at a rate ν_i where we allow for age specific removal rates giving $\nu_i S_i(t)$ as the number of susceptible cattle in the i^{th} age group being removed from the herd at time t. In each simulated epidemic we assume a constant herd size, assuming replacements are drawn from the same age distribution. We make the simpiflying assumption that replacements are all susceptible; in HRAs the effect of having some infectious replacements is subsumed in α while in LRAs, where movements from HRAs are few, the likelihood of multiple introduction is low, due to the low overall prevalence of infection in all cattle (of 5,417,573 tests carried out in 2006 only 20,090 confirmed reactors were found [3]).

As herd size is known to be correlated to infection persistence [4] (Figure 6) we assume density dependent transmission, with infection occuring at at base rate βIS. Exposed cattle become test sensitive at a rate σ and then infectious at a rate γ.

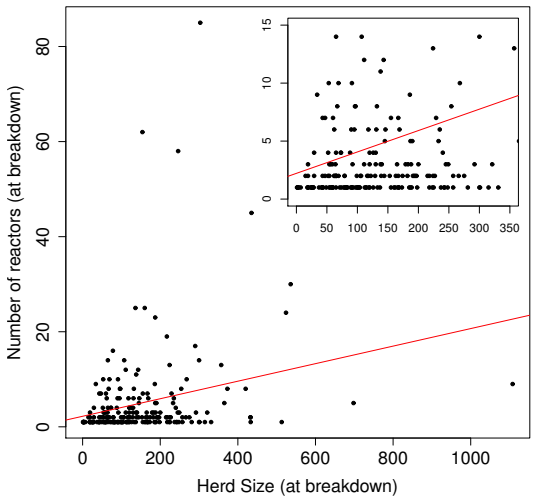


Figure 6: Number of reactors detected to the herd size at breakdown as a function of herd size (inset shows a detailed view of small herds). The correlation is weak but suggests some density dependence (Spearman ρ value of 0.277).

Variability in ζ_H and α are implicit as herd heterogeneity is determined before each individual simulation with fitted probabilities.

We assume outbreaks are initially seeded by a single randomly chosen infected animal. Similar to others [5], we assume that the occurrence of breakdowns due to the introduction of multiple infections is low; this likely causes a compensatory increase in the estimated transmission rate but we expect this effect to be small (see above on national prevalence).

The model (1) was solved by running 2×10^4 independent simulations using Gillespie's τ-leap method with a fixed time step of 14 days to balance simulation efficiency and model accuracy. We used Gillespie Direct Method [6] to validate the choice of time step in the τ-leap method. In each simulation, herd size and age structure are selected from the observed distribution in GB and run to the random predetermined future date selected from a uniform distribution over 0 to n years where n is the testing interval. We then perform a RHT with test sensitivity Ω_r. The number of reactors at the time of a test, N_B, is therefore

$$N_B = binom\left(T(t) + I(t) + I^{super}(t), \Omega_r\right) \tag{2}$$

If no infected cattle are detected we schedule another RHT n years later and continue running the simulation. If, at any time, an animal is removed from the herd it is subjected to a post mortem test with net sensitivity Ω_s, considering the combined probability of being inspected and detected. A positive test triggers a breakdown resulting in a whole herd tuberculin test (WHT). Any confirmed breakdown sets the RHT schedule to every two months until there are no further reactors. When a breakdown is detected (by either RHT or at abattoir), the breakdown size is added to a frequency distribution for routine and triggered WHT tests from which we compare the distribution to the number of reactors at first detection as recorded in VetNet. Each simulation is run for a maximum of 20 years.

The frequency distribution of reactors at first breakdown is interpreted as a multinomial trial with $x_1, x_2, \ldots, x_n$ as the number of times we detected $1, 2, \ldots, n$ reactors in the herd at breakdown in the simulation and we calculate the probabilities $p_1, p_2, \ldots, p_n$ for both the breakdowns detected from routine RHT and abattoir triggered WHT using the observed (age independent) breakdown size distribution. Thus we can write a likelihood function as

$$\mathcal{L} = \frac{n!}{\Pi x_i!} \Pi p_i^{x_i} \tag{3}$$

where n is the total number of breakdowns observed in the simulation and x_i are the frequencies of detecting each breakdown size in our simulations. The free parameters in the described model are $\theta = \{\beta, \sigma, \gamma, \Omega_r, \Omega_s, P_S, P_H, \zeta_S, \zeta_H, \alpha\}$. We calculate the expected value of the model for a given parameter set, θ as

$$E[\theta] = \langle \ell(\mathcal{D}(\theta)|\theta) \rangle$$

where $\mathcal{D}(\theta)$ is the result of simulating the epidemic with θ, and the calculated likelihood is the mean of 10000 simulations of the model as described above. We find that 10000 is the smallest sample size that we can use to have an acceptable variance in our results. We perform a Markov random walk using the Metropolis-Hastings rejection algorithm in the parameter space defined by θ to find the posterior values of θ that correspond the maximum of the likelihood.

We run 5 different chains for each simulation with different initial conditions (θ_0) and run 6000 steps. The posterior distribution is determined by removing the 'burn-in' from the chains. The model was coded in JAVA. The inference scheme was tested for self-consistency by running simulated epidemics using the model structures as defined above and using our inference approach to estimate the model parameters. These were shown to recover the input parameters with posterior distributions of similar width to those identified for our real data, and with the original parameters lying within the 95% credible intervals of the posteriors (results not shown).

Source data

Cattle test data were obtained from the VetNet and Vebus databases obtained from APHA. The cattle distributions throughout GB were obtained from the CTS database (RADAR). We use RHT and WHT records from 2006 because it was sufficiently long after the 2001 foot and mouth disease outbreak for the resultant perturbations in bTB incidence to have disappeared and because quadrennial testing areas expanded rapidly before that year. More recently, the expansion of annual testing areas to cover many regions of relatively low incidence results in additional variability in the testing history of cattle (e.g. many herds moving from 4 year to annual testing in recent years) which would complicate the fitting procedure. This dataset was filtered to consider only breakdowns triggered by RHTs or through tracing from abattoir detection in parishes exclusively tested annually from 1998 to 2006, i.e had a test type of 'VE-RHT', 'VE-WHT', 'VE-WHT2' or 'VE-SLH' in VetNet. Also, we consider only breakdowns where there had not been a previous positive test in the herd and do not consider the results of follow-up tests; this minimizes the impact of possible infections missed from previous outbreaks, and mitigates against possibly epidemiologically significant changes in farmer behaviour after an outbreak begins. This left 315 reactors in herds recorded as beef herds in VetNet, and 208 reactors in dairy herds in high risk. Only herd breakdowns with confirmed reactors (i.e. where visible lesions have been identified) are included; unconfirmed reactors are statistically more likely to show up as reactors later [9]. Inconclusive unconfirmed reactors that never test positive are excluded from the analysis; though there is a statistically significant risk of transmission indirectly associated with them, the absolute risk is slight [10], and therefore likely to be dominated by the transmission from confirmed reactors.

Distribution of priors

The distribution of the number of reactors at first breakdown defines the likelihood function (Equation 3). We calculate the distribution of herd sizes based on the recorded number of cattle when it first suffered a breakdown (i.e. using the same criteria used to obtain the breakdown size distribution to define our likelihood). We use the age structure as found in VetNet, the age distribution of

reactors and the age distribution of cattle sent to slaughter, identifying 14 age groups with lengths from 2-24 months. Uniformly distributed priors based on field and experimental data were used where found (Table 2) [11, 12] with non-informative priors where no estimates existed.

Parameter	Description	Sampling Distribution
β	transmission rate	Uniform$[1 \times 10^{-5}, 1 \times 10^{-2}]$
σ	rate of exposed cattle becoming test sensitive	Uniform[6 hours $-$ 100 days]
γ	rate of test sensitive cattle becoming infectious	Uniform[4 months $-$ 9 months]
Ω_r	probability that a test sensitive or infectious animal is detected by the SICCT test	Uniform$[40 - 80\%]$
Ω_s	probability that a test sensitive or infectious animal is detected at abattoir	Uniform$[50 - 99\%]$
ζ_S	increased infectiousness of superspreaders	Uniform$[1, 1000]$
P_S	proportion of individuals that are superspreaders	Uniform$[1 \times 10^{-3}, 0.4]$
ζ_H	increase of β in herds with high β	Uniform$[1, 1000]$
P_H	proportion of herds with high β	Uniform$[1 \times 10^{-3}, 0.4]$
α	external force of infection	Uniform$[1 \times 10^{-10}, 5 \times 10^{-3}]$

Table 2: Summary of the priors used in the model.

For $\sigma, \gamma, \Omega_r, \Omega_s$ the priors were chosen on the basis of existing field and experimental estimates [12] [11], non-informative priors were used for all other parameters. The rate exposed cattle become test sensitive and test sensitive cattle become infectious is the inverse of the exposed and test sensitive periods (6hours $-$ 100days and 4months $-$ 9months) respectively.

Results

1. **Development and analysis of locations of HRAs and TAs**

Identification of HRAs

The sequential aggregation of ranked quadrats into putative set of HRAs shows an increased model likelihood with no fluctuations. Figure 7 shows that the implementation of the new ordering method eliminated the fluctuations originated by the old method of aggregating quadrats (blue and red lines). The green line shows that just one iteration is necessary to correctly order the quadrats showing a decrease in accuracy when implemented a series of three iterations.

Beyond the maximum-likelihood value, likelihoods typically declined at a lower rate per quadrat, compared to the initial increase in likelihood up to the maximum. This shows that, while HRAs are very well defined, there is higher uncertainty in identifying what areas are not HRA, possibly because these usually contain fewer breakdowns due to an inherit lower incidence and because these are typically in quadrennial testing areas.

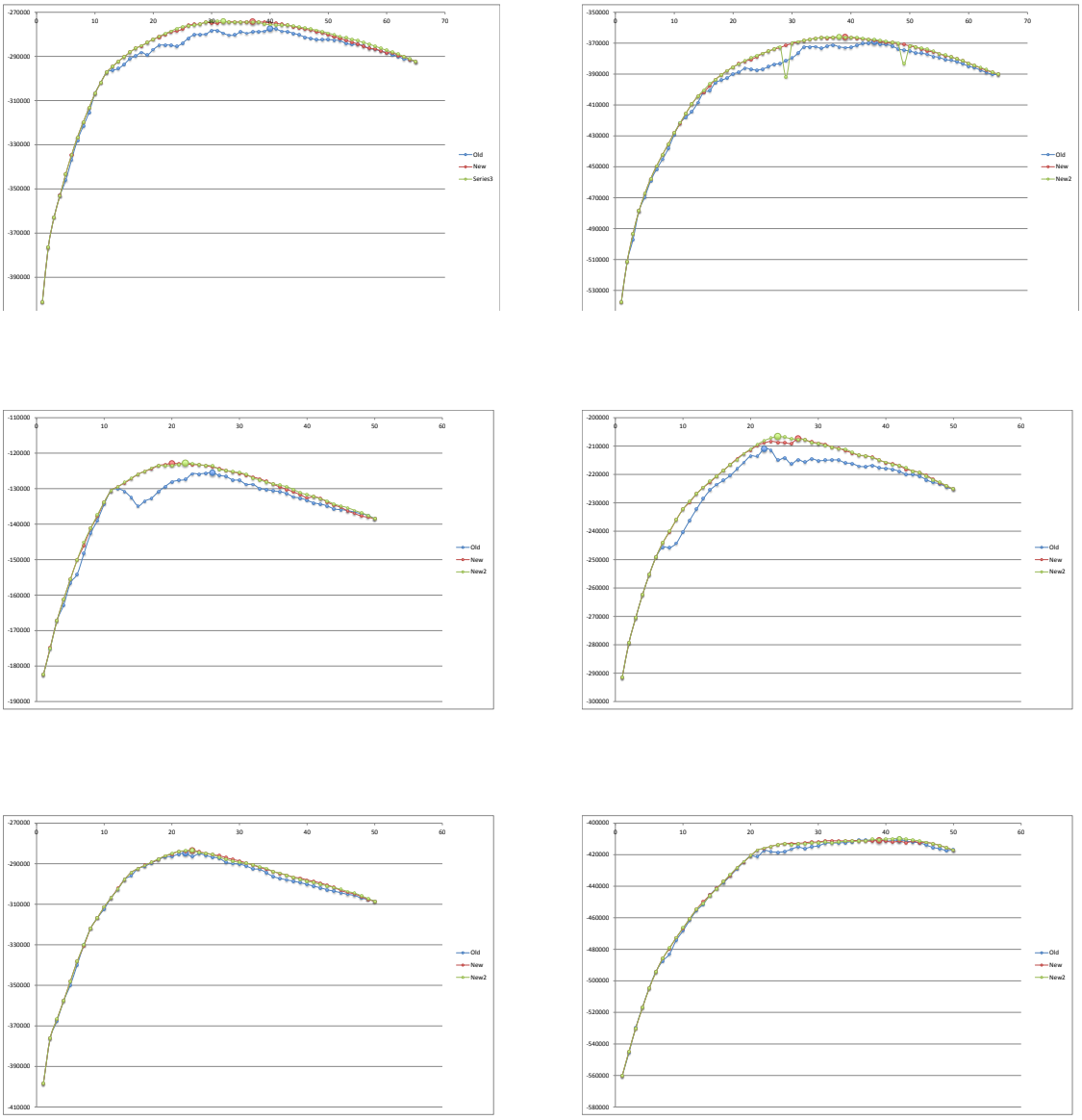


Figure 7: Aggregated likelihood statistics for the three different genotypes: S9:b (top), S10:a (middle) and S25:a (bottom). The different lines represent 1) the older method of aggregating the quadrats (blue line: old); 2) the newer method of aggregating the quadrats (red line: new), and 3) a repetition of three series of the newer method (green: series3). Results for the years 2003 (left) and 2007 (right) are shown for the different genotypes.

Identification of TAs

Transitional areas were defined for each spoligotype as quadrats that that have been intermittently high risk during the period 2003-2007 and that are high risk in 2007. The map in Figure 8 shows the spatial position of herds located in the areas identified as transitional for all the three spoligotypes.

Figure 8: Map showing herds located in transitional areas for the spoligotypes S9b, S10a, and S25a. A transitional area for each spoligotype was defined as the set of quadrats that became infected or became cleared of infection during the study period 2003-2007.

2. **Statistical analysis of underlying risk factors for breakdowns in areas of different risk**

   The continuing expansion of high-incidence areas of bTB in GB raises a number of questions concerning the determinants of infection at the herd level that are driving the spread of the disease. Here, we develop risk factor models to understand which risk factors can better predict bTB breakdowns, and to quantify the importance of different risk factors, such as herd sizes, herd type, imputed badger densities and cattle movements.

2.1. Risk factors as predictors of bTB breakdowns in Scotland and LRA England

In order to understand if movements from HRAs are good predictors of bTB breakdowns in Scotland and in LRA England and how these areas differ from each other, we ran different risk ratings for herds as stump classifiers for different thresholds (as described in Methods). These were compared with bTB breakdowns and we recorded the number of true and false positives.

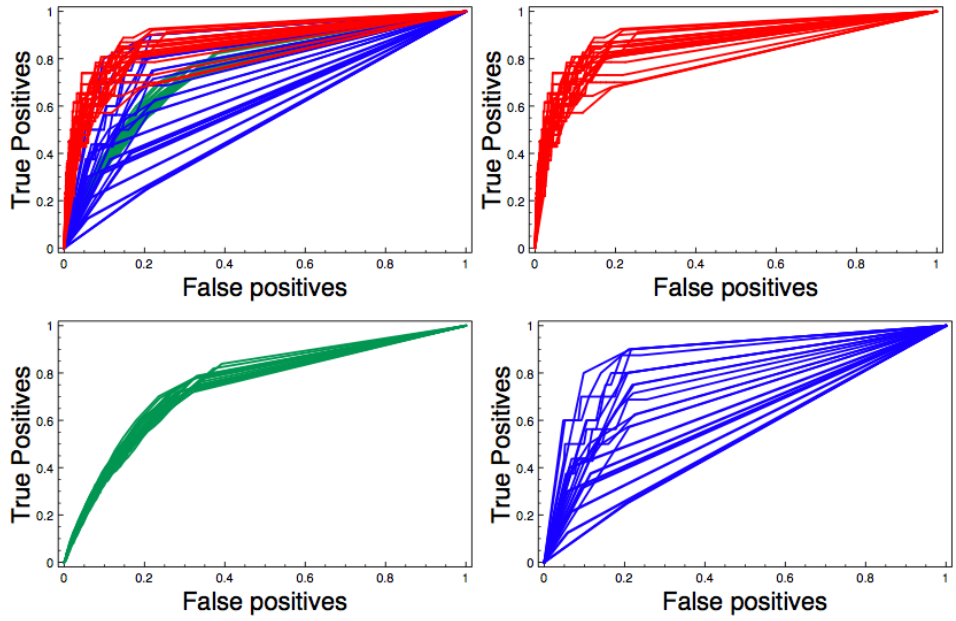


Figure 9: ROC curves from all runs of the stump classifier in LRA England (red), Scotland (blue) and the entire United Kingdom (green).

Figure 9 shows that by incorporating risk ratings in our analyses, our prediction of breakdowns was improved in the three areas of study. Since LRA England curves are more concentrated on the upper left side of the graph, our prediction of breakdowns in LRA England using movements as risk factors is better than in Scotland or in the rest of the UK.

2.2. Statistical analysis of underlying risk factors for breakdowns in LRAs, HRAs and TAs

The use of stump classifiers allow us to rapidly evaluate the role of different movement types as risk factors for spread, and broadly speaking showed us that movements from HRAs to LRAs in England were a good (better than random) predictor of risk. In order to compare to the previous work in analysing risk factors in Scotland, we then used a regression-based statistical analysis, as in [12] for a more detailed analysis including non-movement risk factors.

We compared the significance of the different risk factors in LRAs, HRAs and TAs (as determined by long term parish testing interval consistency for LRAs and HRAs and by simulation results for TAs, respectively).

Predictor	Unit	Odds (95% CIs)	Z value	Pr(>|Z|)
Intercept			3.948	< 0.001
x-coordinate	x/100000	0.010 (0.003, 0.040)	-6.565	< 0.001
y-coordinate	y/100000	0.094 (0.046, 0.191)	-6.536	< 0.001
Herd type	Other	1		
	Fattening	2.127 (1.293, 3.501)	2.971	< 0.001
Herd Size	0-9	1		
	10-99	0.839 (0.367, 1.921)	-0.414	0.346
	≥ 100	3.445 (1.749, 6.784)	3.577	< 0.001
Movements from HRAs	0	1		
	1-10	1.407 (0.883, 2.243)	1.436	0.145
	> 10	4.203 (2.503, 7.058)	5.430	< 0.001
Irish imports	No	1		
	Yes	6.248 (4.133, 9.445)	8.691	< 0.001
x*y		1.851 (1.518, 2.258)	6.075	< 0.001

Table 3: Risk factor model for Scotland during the period between 2002 and 2008.

Predictor	Unit	Odds (95% CIs)	Z value	Pr(>|Z|)
Intercept		0 (0.000, 0.000)	-27.292	<0.000
Herd type	Beef	1		
	Dairy	0.577 (0.359, 0.926)	-2.277	0.023
	Stores	1.061 (0.621, 1.810)	0.216	0.829
	Suckler	0.423 (0.264, 0.679)	-3.565	<0.001
	Finishing	1.263 (0.851, 1.876)	1.158	0.247
Herd size	1-50	1		
	51-100	2.813 (1.594, 4.962)	3.570	<0.001
	101-200	4.237 (2.484, 7.225)	5.302	<0.001
	201-350	4.489 (2.497, 8070)	5.019	<0.001
	351-500	10.238 (5.500, 19.057)	7.337	<0.001
	>500	11.509 (6.324, 20.945)	7.997	<0.001
Batches from HRAs	0	1		
	1-10	3.586 (2.555, 5.032)	7.387	<0.001
	> 10	12.444 (7.983, 19.399)	11.130	<0.001

Table 4: Risk factor model for LRA in England during the period between 2002 and 2008.

Risk factors for LRAs in England in 2002-2008 are presented in Table 4. Of the variables considered in this analysis, herd type, herd size and batches from HRAs were shown to have any significance. The risk of detecting a bTB infection in dairy, stores and finishing herds was not significantly different from beef herds, while the results suggest that suckler herds have a protective effect (Odds: 0.423). Herds larger than 50 were at greater risk than those under 50, with herds with 51-100 animals being 2.81 times more likely to have a breakdown. This increased to 4.24, 4.49, 10.24 and 11.50 for 101-200, 201-350, 351-500 and > 500 respectively. The number of batches from HRAs were considered a risk, where herds that received between 1 and 10 batches per year are 3.59 times more likely to have a breakdown than the ones that did not receive animals from high risk areas. This value increased to 12.44 when herds import more than 10 batches a year. This result is consistent with the results from the stump classifier analysis (Figure 9), where cattle movements were considered a good predictor of risk.

Risk factors for LRAs of England in 2008-2013 are presented in Table 5. Of the variables considered in this analysis, herd type, herd size, number of batches from HRAs, imports from Ireland and interactions between imputed badger density and herd type were shown to have any significance. An interaction between herd type and imputed badger density was significant, meaning that a beef herd in a badger high density area was at increased risk for bTB (Odds=3.57), while dairy herds in the same area were less likely to contract the infection (Odds=0.10). The remaining interactions with imputed badger density and herd type were shown not significant. Similar to the earlier time period, herds larger than 100 animals were increasingly, and more dramatically, associated with having confirmed bTB reactors. Herds with 101-200, 201-350, 351-500 and >500 were 6.04, 9.68, 15.01 and 20.48 times more likely to have a breakdown, respectively. While still above the threshold of statistical significant, the associated confidence intervals suggest a great deal of uncertainty with the aforementioned Odds.

Predictor	Unit	Odds (95% CIs)	Z value	Pr(>|Z|)
Intercept	0	0	-20.680	<0.001
Herd type	Beef	1		
	Dairy	0.830 (0.397, 1.732)	-0.497	0.619
	Stores	1.571 (0.676, 3.653)	1.050	0.294
	Suckler	1.578 (0.842, 2.956)	1.423	0.155
	Finishing	1.883 (0.978, 3.627)	1.893	0.058
Badger density	Rescaled	3.566 (0.913, 13.932)	1.829	0.067
Herd size	1-50	1		
	51-100	3.952 (1.800, 8.676)	3.425	0.001
	101-200	6.042 (2.866, 12.735)	4.728	<0.001
	201-350	9.680 (4.523, 20.717)	5.848	<0.001
	351-500	15.014 (6.617, 34.065)	6.481	<0.001
	>500	20.476 (9.415, 44.530)	7.617	<0.001
Batches from HRAs	1	0		
	1-10	2.837 (1.864, 4.317)	4.868	<0.001
	> 10	11.985 (7.218, 19.899)	9.601	<0.001
Imports from Ireland	False	1		
	True	2.253 (1.306, 3.887)	2.918	0.004
Interactions	Badger-Beef	1		
	Badger-Dairy	0.099 (0.020, 0.498)	-2.806	0.005
	Badger-Finishing	0.791 (0.153, 4.084)	-0.280	0.779
	Badger-Stores	0.769 (0.092, 6.458)	-0.242	0.809
	Badger-Suckler	0.299 (0.066, 1.344)	-1.575	0.115

Table 5: Risk factor model for LRA in England and Wales during the period between 2008 and 2013.

Risk factors for HRAs in England and Wales in 2002-2008 are presented in Table 6. Unlike LRAs during the same time period, herd type and herd size were the only significant risk factors

in predicting confirmed bTB reactors. Only dairy and finishing herds were more risky when compared to beef herds (Odds=0.99 and 1.96, respectively), while stores and suckler herds were not significantly different from beef herds. Herds with more than 50 animals where shown to be more risky than herds with less than 50 animals, where more than 500 animals were shown to be only 6.55 times more risky, though this is not significantly different from the Odds corresponding to the 351-500 animals category.

Predictor	Unit	Odds (95% CIs)	Z value	Pr(>|Z|)
Intercept		0.045 (0.037, 0.055)	-31.826	<0.000
Herd type	Beef	1		
	Dairy	0.989 (0.919, 1.065)	-0.281	<0.001
	Stores	0.904 (0.800, 1.020)	-1.637	0.102
	Suckler	0.981 (0.911, 1.057)	-0.497	0.619
	Finishing	1.965 (1.345, 2.872)	3.490	<0.001
Herd size	1-50	1		
	51-100	2.530 (2.330, 2.746)	22.194	<0.001
	101-200	3.933 (3.641, 4.249)	34.787	<0.001
	201-350	5.691 (5.231, 6.191)	40.472	<0.001
	351-500	7.239 (6.479, 8.089)	34.978	<0.001
	>500	6.548 (5.824, 7.362)	31.438	<0.001

Table 6: Risk factor model for HRA in England during the period between 2002 and 2008.

Risk factors for HRAs in England and Wales in 2008-2013 are presented in Table 7. Herd type, herd size and number of batches from high risk areas were the only significant risk factors in predicting confirmed bTB reactors. Similar to the earlier time period in HRAs, finishing herds were significantly associated with an increase in a breakdown (Odds=1.17). Increases in herd size over 50 were associated with increased risk, where more than 500 animals were shown to be only 5.82 times more risky, though this is not significantly different from the Odds corresponding to the 351-500 animals category (just like in the previous study period). Herds with 51-100, 101-200, 201-350, 351-500 and over 500 animals were 2.52, 3.78, 5.50, 7.02 and 6.36 times more likely to have a breakdown when compared to herds with 1-50 animals. The results related to cattle movements suggest that receiving animals from herds in already high risk areas does not increase the risk of bTB (Odds=0.54).

Predictor	Unit	Odds (95% CIs)	Z value	Pr(> |Z|)
Intercept		0.030 (0.023, 0.040)	-24.518	<0.001
Herd type	Beef	1		
	Dairy	1.001 (0.943, 1.062)	0.036	0.971
	Stores	0.992 (0.905, 1.088)	-0.166	0.868
	Suckler	1.017 (0.958, 1.079)	0.548	0.583
	Finishing	1.166 (1.091, 1.247)	4.499	<0.001
Herd size	1-50	1		
	51-100	2.523 (2.378, 2.677)	30.592	<0.001
	101-200	3.785 (3.577, 4.004)	46.338	<0.001
	201-350	5.495 (5.163, 5.847)	53.678	<0.001
	351-500	7.016 (6.479, 7.597)	47.986	<0.001
	>500	6.364 (5.875, 6.892)	45.433	<0.001
Batches from HRAs	0	1		
	1-10	0.540 (0.450, 0.647)	-6.652	<0.001
	> 10	0.735 (0.547, 0.988)	-2.042	0.041

Table 7: Risk factor model for HRA in England between 2008 and 2013.

The risk factors for the TAs are presented in Table 8.

Predictor	Unit	Odds (95% CIs)	Z value	Pr(> |Z|)
Intercept		0.014 (0.008, 0.024)	-15.7798	<0.001
Herd size	1-50	1		
	51-100	2.989 (2.368, 3.775)	9.205	<0.001
	101-200	5.066 (4.097, 6.264)	14.978	<0.001
	201-350	8.179 (6.558, 10.202)	18.640	<0.001
	351-500	10.015 (97.533, 13.3150)	15.855	<0.001
	>500	10.778 (7.946, 14.619)	15.287	<0.001
Batches from HRAs	0	1		
	1-10	0.627 (0.528, 0.744)	-5.326	<0.001
	> 10	0.732 (0.516, 1.038)	-1.753	0.080

Table 8: Risk factor model for TA in England between 2003 and 2007.

Herds with 51-100 animals were 2.99 times more likely to have a breakdown, with 101-200 animals and 201-350 animals being 5.07 and 8.18 times more risky. This value increased to 10.02 times the risk of herds with 351-500 animals when compared to small herds 1-50 animals. herds with >500 animals (Odds=10.78) do not seem to be significantly more risky than herds with 351-500 animals. Just like in HRAs for the study period 2008-2013, the number of batches received from high risk areas do not seem to increase the risk of infection.

Each risk factor model included herd type (except for TAs) and herd size as significant risk factors for herd level prevalence of *M. bovis* infection, however, no particular herd type was consistently risky. This may reflect differences in the sample size of each farm type in the study areas as opposed to regional variation in the riskiness of farming practices for different herd types. Previous risk factor assessments in LRAs of Great Britain have highlighted movements from HRAs of England and Wales as well as imports from Northern and the Republic of Ireland ([19]). Movements from high risk areas were identified as significant risk factors in low risk areas for both study periods, in the later period of high risk areas and in transitional areas. While in low risk areas, these increase the risk of getting infected, in high risk areas it seems that they are not relevant, since the herds are already at a higher risk of infection. The influence of cattle movements from high risk areas into transitional areas also seems to play a very minor role. More research on the expansion of transitional areas needs to be done to better understand the causes associated with spatial increase of infection. The discrepancy of the identified risk factors across the different incidence areas may be a product of the more conservative model fitting selection criteria or possible relationships between the different factors that can mask the contribution of movements from HRAS.

The statistical results suggesting that imputed badger density is a significant risk factor in low incidence areas but not in high incidence areas bears some comment given the known epidemiology of the disease. The herd-level prevalence of infection in low incidence areas calculated from the model considering estimated main badger sett density exclusively suggests geographical clustering of areas with high estimated density of setts, with areas adjacent to the high incidence areas more likely to contract the infection. It is therefore possible that this risk factor may be a proxy for other geographically localised correlates. Figures 10, 11, and 12 show how badger sett density (provided by APHA) is related with herd size in areas of different risk (LRAs and HRAs) and for two different study periods (2003-2008 and 2009-2013).

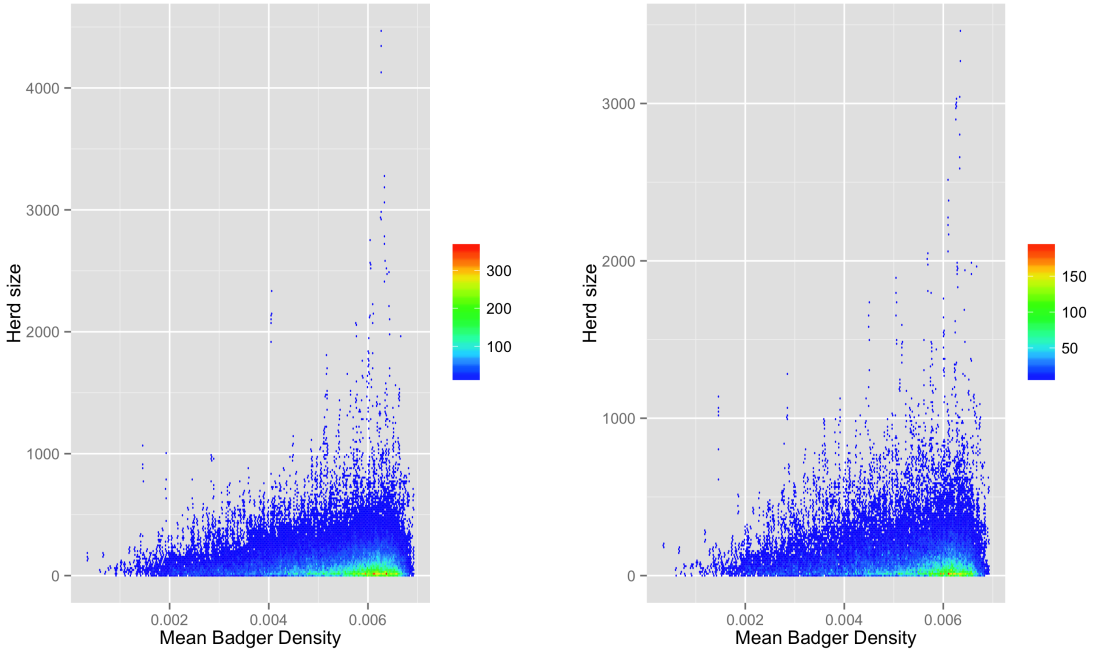


Figure 10: Smooth scatter plots of the mean imputed badger density versus herd sizes for LRA England during the periods 2003-2008 and 2009-2013.

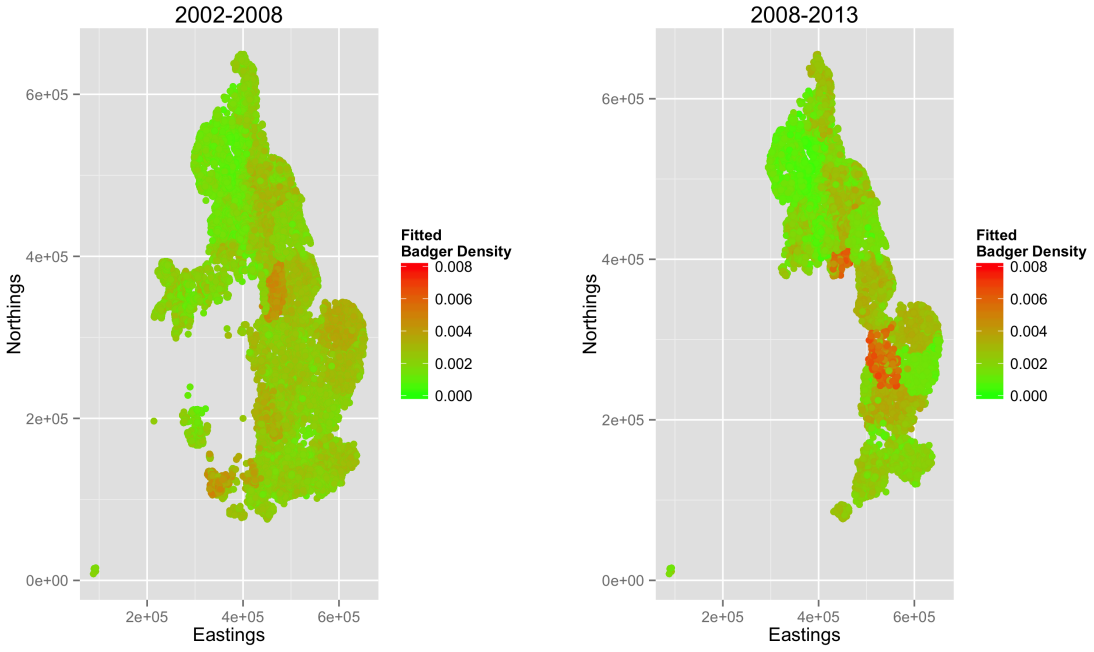


Figure 11: Fitted value for imputed badger density in mixed regression model for LRA England during the periods 2003-2008 and 2009-2013.

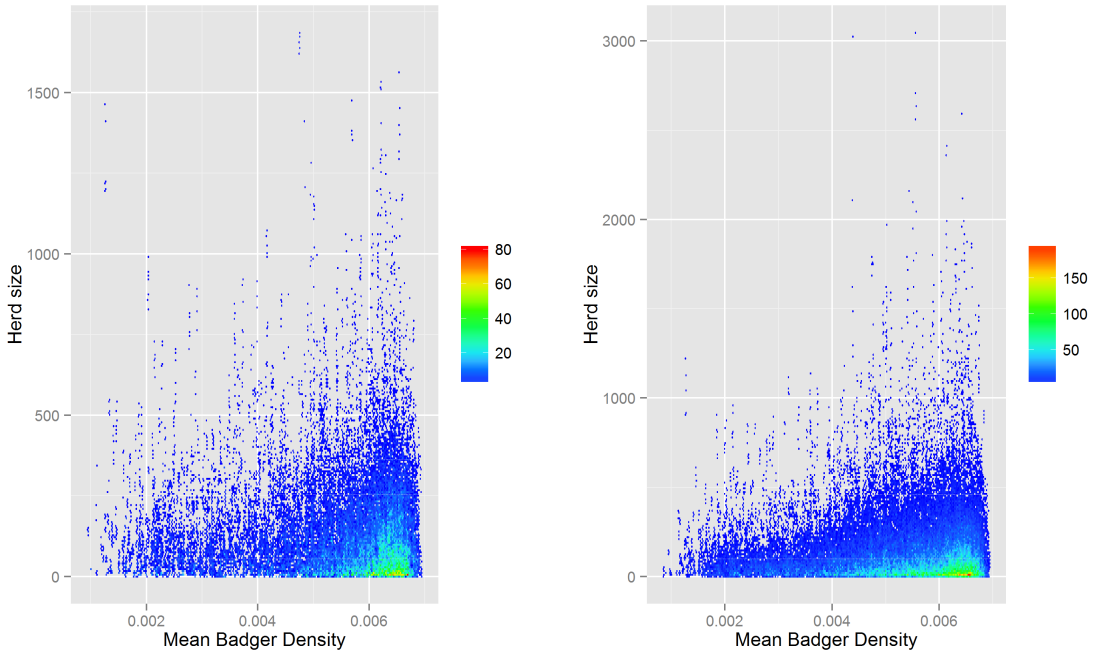


Figure 12: Smooth scatter plots of the mean imputed badger density versus herd sizes for HRA England during the periods 2003-2008 and 2009-2013.

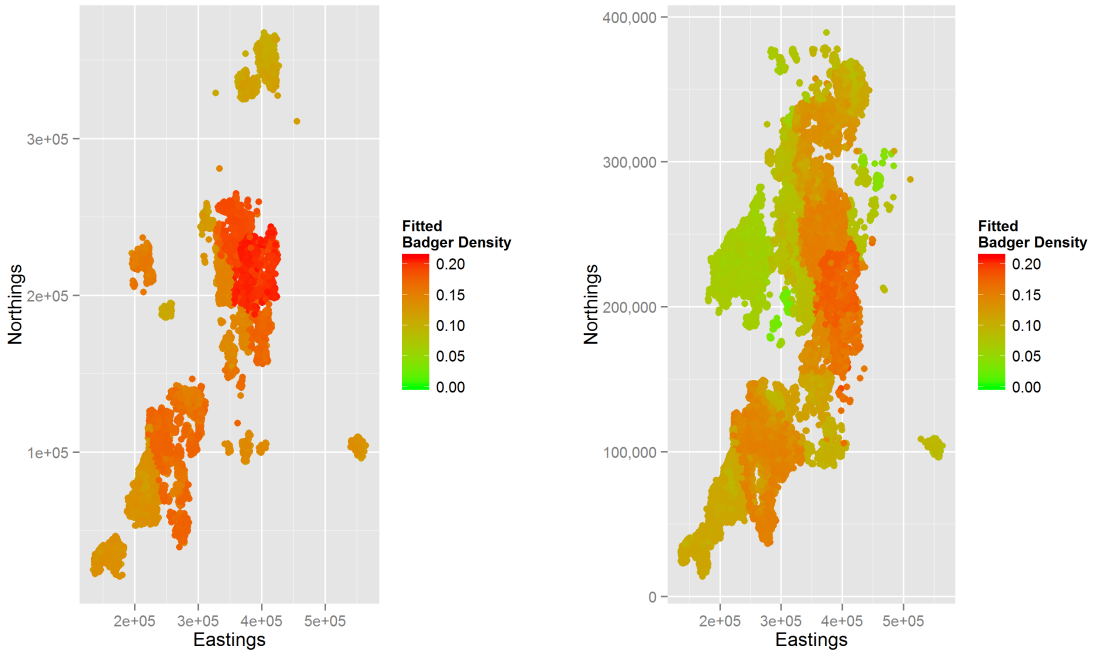


Figure 13: Fitted value for imputed badger density in mixed regression model for HRA England during the periods 2003-2008 and 2009-2013.

Further, in order to facilitate fitting estimated main badger sett density to the models as described above, the data required normalisation, a process carried out on both the 2002-2008 and 2008-2013 low and high incidence datasets individually. It is possible that this transformation could accentuate the differences in estimated density values, giving the impression of a greater range then the original national data suggest and overemphasising any statistical relationship found. Complications with interpreting the results may also arise from fitting rescaled imputed badger

density to a model without consideration of the practical range in values. The inclusion of imputed badger density and herd type interactions in the 2008-2013 low incidence risk factor analysis may nevertheless be indicative of a relationship masked in the 2002-2008 dataset. However, even though the earlier dataset is much larger and more geographically diverse, a further investigation into the relationship between the different herds and areas of high imputed badger density may help in understanding a potentially emerging risk factor. This outcome, while resulting in a model that best fits the data, may be merely a result of correlates to imputed badger density estimation (detailed above) and imputed badger density transformation to permit model fitting, rather than density of badgers themselves.

3. **Application of the Scottish risk-based surveillance model to long term quadrennial testing areas to reduce testing.** Currently, Scotland has implemented a surveillance system based on work by Bessell *et al.* [12] and all eligible herds in LRA England are tested once every 4 years under the routine herd testing (RHT) policy. RHT is supplemented by further whole herd tests, triggered by post-movement testing of cattle moving from HRA England, from Wales or Ireland and contact tracing following a confirmed breakdown for Scotland (2003-2008) (Table 9), for LRA England (2003-2008) (Table 10), and for LRA England (2009-2013) (Table11). The type of tests shown for Scotland are similar to the ones presented by Bessell *et al.* [12], while we have explored a few more for LRAs Englands shown in Tables 10 and 11. These show that most confirmed breakdowns have been detected due to slaughterhouse, tracing and WHT tests (either directly or indirectly through inconclusive reactor retests).

Test type	Part/Whole herd	Herds tested	Confirmed breakdowns	Animals tested	Reactors	Confirmed reactors
RHT	Whole	20580	21	1241389	46	22
Tracing	Part	2984	18	8393	76	22
Post-movement	Part	1650	1	13367	2	0
Post-Irish import	Part	1077	9	16402	31	11
Inconclusive reactor retest	Part	1530	47	2544	267	54

Table 9: The number of herds and animals tested by different surveillance types in Scotland between 2003 and 2008.

Test type	Part/Whole herd	Herds tested	Confirmed breakdowns	Animals tested	Reactors	Confirmed reactors
Contiguous herd	Whole	361	1	37480	3	1
1st hotspot check	Whole	35	0	4004	0	0
Pre-Movement	Part	1271	0	17780	0	0
RHT	Whole	21907	18	1120146	364	31
WHT	Whole	1032	1	82522	1	1
Tracing	Part	9095	41	30638	72	54
Post-movement	Part	275	0	1062	0	0
Post-Irish import	Part	169	0	719	0	0
Inconclusive reactor retest	Part	2036	19	4909	52	18
Slaughterhouse	Part	110	96	0	323	162

Table 10: The number of herds and animals tested by different surveillance types in LRA England between 2003 and 2008.

Test type	Part/Whole herd	Herds tested	Confirmed breakdowns	Animals tested	Reactors	Confirmed reactors
Contiguous herd	Whole	420	1	49459	1	1
1st hotspot check	Whole	24	0	3269	0	0
Pre-Movement	Part	1899	0	26499	3	3
RHT	Whole	6069	3	326669	7	4
WHT	Whole	697	2	45872	6	2
Tracing	Part	6007	36	21182	59	43
Post-movement	Part	277	1	972	0	0
Post-Irish import	Part	280	4	3321	15	3
Inconclusive reactor retest	Part	651	23	1279	78	15
Slaughterhouse	Part	163	71	0	104	89

Table 11: The number of herds and animals tested by different surveillance types in LRA England between 2009 and 2013.

Analysis of the squared root of the mean number of animals sent to slaughter per year during the periods 2003-2008 (in Scotland and LRA England), and 2009-2013 (in LRA England) against the squared root of the mean herd size shows a linear relationship with distinct clustering of finishing and dairy herds (Figure 14). Finishing herds are typically smaller than dairy herds, however, they have more *per capita* movements to slaughter compared to dairy herds.

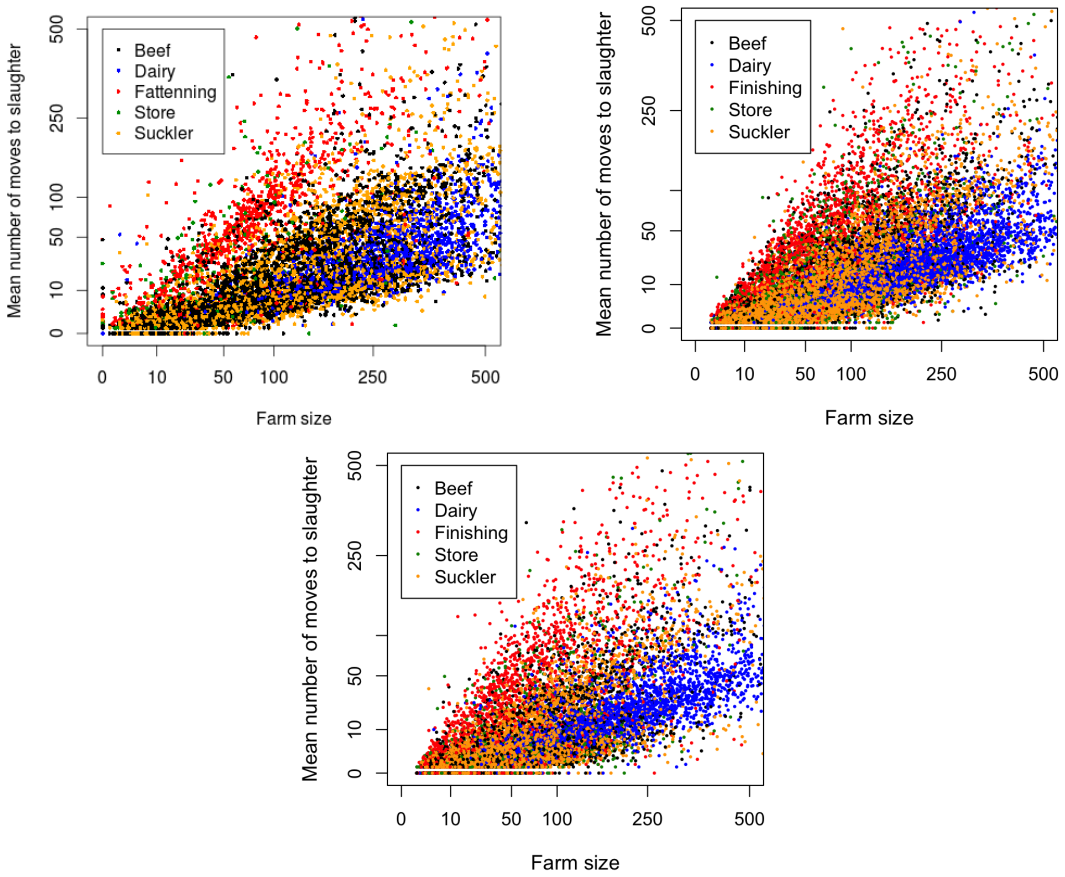


Figure 14: Scatterplots showing the squared root of mean herd size per herd against the squared root of the mean number of animals moved to slaughter in Scotland (top left), LRA England (top right) during 2003-2008, and LRA England (bottom) during 2009-2013 broken down by herd type as listed in SAM's database. The axes have been truncated for clarity. There is a linear relationship between herd size and number of movements to slaughter with distinct clustering of finishing and dairy herds. Albeit dairy herds are larger, they have fewer per capita movements to slaughter compared to fattening herds.

For a comparison of alternative surveillance scenarios, we applied the scenarios defined in [12] (based upon identifying herds that are more likely to be infected and not detected) to Scotland during 2003-2008, and we applied the eight scenarios defined in the Methods section to the LRA England during the time periods 2003-2008 and 2009-2013. A multi-based risk is examined hereafter. Results of the four baseline scenarios - *Slaughterhouse only*, *Four year*, *Two year* and *One year* surveillance models applied to Scotland and LRA England during 2003-2008 and to LRA England during 2009-2013 are presented in Figures 15 and summarized in Tables 12, 13, and 14.

Figure 15 shows the cumulative distribution plots of the probability of each herd harbouring undetected infection at the end of each model time step for Scotland 2003-2008 (top left), LRA England 2003-2008 (top right), and LRA England 2009-2013 (bottom). There is a small percentage of herds that have a very low probability of undetected infection, and that the probability of undetected infection varies with the different testing regimes adopted. As expected, *slaughterhouse* surveillance allows for the largest proportion of herds to have a higher probability of undetected infection, followed by the *four-year* and all the risk-based scenarios, and lastly by the *one-year* and *two-year* testing scenarios, in which large proportion of herds have at most a very low probability of undetected infection. As an example, with *slaughterhouse* surveillance only, around 90% of the herds have a probability of undetected infection of at most 0.011, 0.010 and 0.012 for, respectively Scotland, LRA England 2003-2008 and England 2009-2013; with *four-year* testing, the same percentage of herds have a probability lower than 0.004 per herd in all three areas, while if the frequency of testing is increased in England for *one-year* or *two-year* testing, the probability is at most 0.003 for *two-year* testing and at most 0.001 for *one-year* testing for both study periods in England.

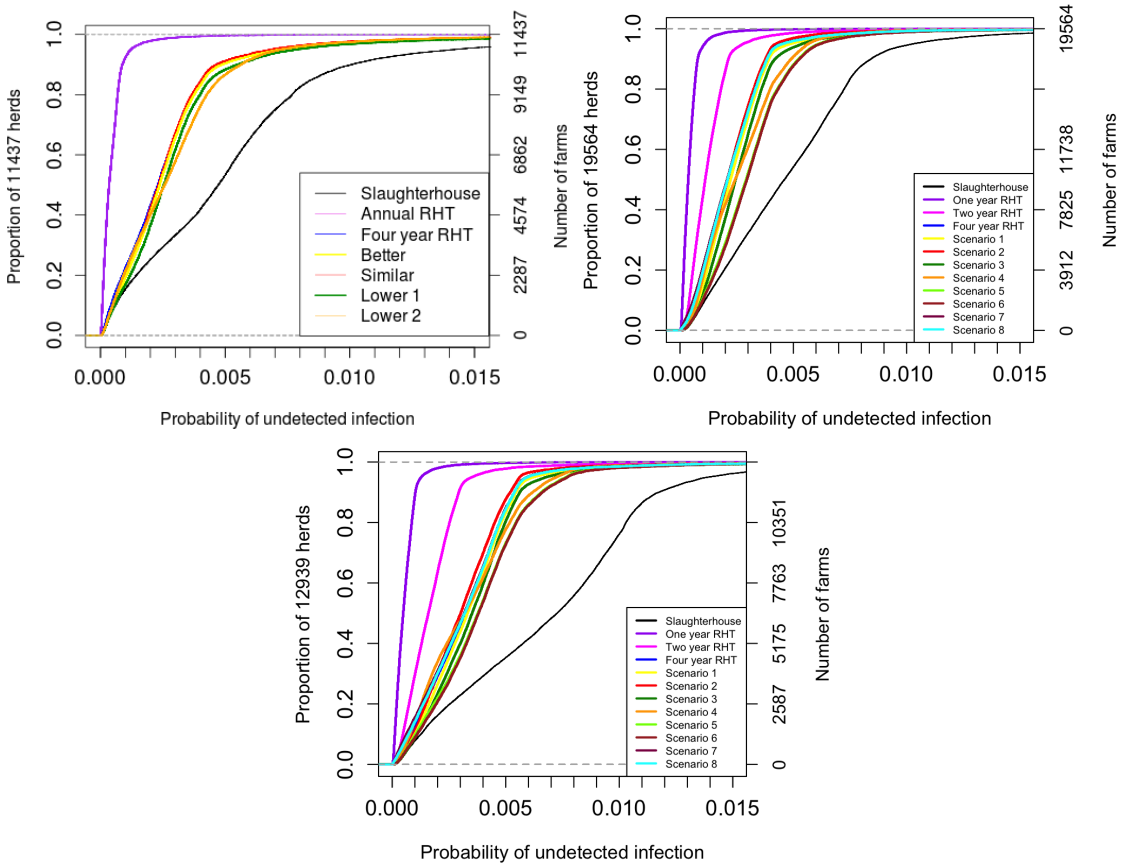


Figure 15: Cumulative distribution plot of the probability of each herd harbouring undetected infection at the end of each model time step for all herds in Scotland (top left) and in LRA England (top right) during 2003-2008, and LRA England during 2009-2013 (bottom) from the three baseline models. The 'Slaughterhouse only' scenario represents a minimal model (the lowest amount of surveillance that could be under) and the 'Annual routine herd testing' scenario is the maximal model (the most surveillance that could be undertaken under the current testing regimens). The remaining four (top left)) and eight scenarios (top right) and bottom)) are the four- and eight- risk based surveillance systems used, respectively in Scotland and in LRA England. The x-axis has been truncated for clarity.

Regarding the risk-based surveillance scenarios, 90% of the herds will have a probability between 0.004 and 0.006, in which the *better* and *similar* scenarios have a lower probability than the *lower* detection scenarios (Scotland 2003-2008). For LRA England 2003-2008, 90% of the herds have at most probability between 0.004 and 0.006 per herd for all eight surveillance scenarios. *Scenarios 1-2* and 7-8 have very similar curves and have lower probabilities of undetected infection per herd than *Scenarios 5-6*. These patterns hold for all extra scenarios in LRA England 09-13, however, the probability values vary between 0.004 and 0.007, suggesting that there is a slight increase in probability of undetected infection per herd over time.

Surveillance scenario	Interval (years)	Herds tested	Cattle tested	Fitted no. detected infections	Latent infections		False positives 2008
					2008	Mean	
Baseline scenarios							
Slaughterhouse only	n.a	0	0	86.70	54.54	48.26	0
Current	**4**	**2859**	**438854**	**106.47**	**20.51**	**21.27**	**62.96**
Maximum	1	11437	1755418	107.20	2.72	2.95	253.95
Risk-based surveillance							
Better	2/4	2292	352454	107.36	19.99	21.00	51.10
Similar	4	2138	309040	106.08	21.40	22.06	46.55
Lower Detection 1	4	1667	200405	104.30	23.88	24.03	34.82
Lower Detection 2	1/2/4	1912	396516	104.18	24.51	24.56	46.48

Table 12: Summary of test results for Scotland (2003-2008).

Surveillance scenario	Interval (years)	Herds tested	Cattle tested	Fitted no. detected infections	Latent infections mean	Latent infections 2008	False positives mean	False positives 2008
Baseline scenarios								
Slaughterhouse	n.a	0	0	139.83	76.50	84.57	0.00	0.00
Four year	**4**	**4891**	**554723**	**160.96**	**30.81**	**30.71**	**107.95**	**105.19**
Two year	2	9782	1109446	158.73	15.02	15.48	215.78	213.30
One year	1	19564	2218893	153.96	4.69	5.22	432.62	422.07
Risk-based surveillance								
Scenario 1	2/4	4262	490551	163.30	29.54	28.80	94.07	90.77
Scenario 2	4	3729	408355	160.41	32.26	32.35	81.47	80.53
Scenario 3	4	2823	287405	157.95	35.73	36.49	61.20	60.44
Scenario 4	1/2/4	3301	534062	155.75	38.06	39.37	80.75	79.98
Scenario 5	1/2/4	1811	271122	150.55	44.17	47.29	40.90	41.53
Scenario 6	1/2/4	1654	205482	149.96	44.68	48.03	35.39	33.82
Scenario 7	1/4	5427	765500	161.58	30.02	29.71	124.85	123.03
Scenario 8	1/4	5173	732626	161.44	30.34	30.12	117.55	115.33

Table 13: Summary of test results for LRA England (2003-2008).

Surveillance scenario	Interval (years)	Herds tested	Cattle tested	Fitted no. detected infections	Latent infections mean	Latent infections 2013	False positives mean	False positives 2013
Baseline scenarios								
Slaughterhouse	n.a	0	0	116.85	85.52	91.11	0.00	0.00
Four year	**4**	**3234**	**391590**	**145.15**	**30.89**	**26.89**	**71.41**	**69.44**
Two year	2	6469	783181	142.83	14.52	12.06	143.24	138.69
One year	1	12939	1566363	136.70	4.20	3.52	286.86	279.15
Risk-based surveillance								
Scenario 1	2/4	3064	369651	147.61	28.42	24.08	67.88	66.26
Scenario 2	4	2531	290557	144.72	32.01	28.03	55.63	55.03
Scenario 3	4	2074	220600	143.25	34.32	30.61	44.95	43.46
Scenario 4	1/2/4	2827	448114	143.47	32.92	29.70	68.49	67.59
Scenario 5	1/2/4	1803	264069	140.03	38.35	35.73	40.92	39.63
Scenario 6	1/2/4	1682	213142	139.62	38.90	36.33	36.72	35.75
Scenario 7	1/4	3626	556163	145.70	29.95	25.89	84.21	82.50
Scenario 8	1/4	3486	570642	145.44	30.41	26.42	79.58	77.94

Table 14: Summary of test results for LRA England (2009-2013).

According to Tables 13 and 14, *slaughterhouse* surveillance alone detects fewer breakdowns than the *four-year* baseline scenario (140 compared to 161 in 2003-2008; 117 compared to 144 in 2009-2013) and presents a higher number of latent infections per year (76 compared to 31 in 2003-2008; 86 compared to 31 in 2009-2013). The *two-year* baseline scenario detects a similar number of breakdowns as the *four-year* one (159 compared to 161 in 2003-2008; 143 compared to 145 in 2009-2013), fewer latent infections per year (15 compared to 31 in 2003-2008; 14 compared to 31 in 2009-2013), but presents more false positives (216 compared to 108 false positives in 2003-2008; 143 compared to 71 false positives in 2009-2013), and more herds tested (9782 compared to 4891 herds tested in 2003-2008; 6469 compared to 3234 herds tested in 2009-2013). The *one-year* baseline scenario detects slightly fewer breakdowns than the *four-year* one (154 compared to 161 in 2003-2008; 137 compared to 145 in 2009-2013), fewer latent infections per year (5 compared to 31 in 2003-2008; 4 compared to 27 in 2009-2013), but with many more false positives (433 compared to 108 false positives in 2003-2008; 287 compared to 71 false positives in 2009-2013) and a substantial increase of herds tested (19564 compared to 4891 in 2003-2008; 12939 compared to 3234 herds tested in 2009-2013).

Regarding the different risk-based surveillance scenarios, *Scenarios 1-2* detect a very similar number of breakdowns as the *four-year* regime (163 and 160 compared to 161 in 2003-2008; 144 and 143 compared to 144 in 2009-2013), a very similar number of latent infections per year (30 and 32 compared to 31 in 2003-2008; 36 and 37 compared to 35 in 2009-2013), and a fewer number of false positives (94 and 81 compared to 108 in 2003-2008; 59 and 55 compared to 73 in 2009-2013), for fewer herds tested (4262 and 3729 compared to 4891 in 2003-2008; 2743 and 2549 compared to 3347 in 2009-2013). *Scenarios 3-6* present similar patterns between them: fewer herds tested than the *four-year* baseline scenario (2823, 3301, 1811 and 1654 compared to 4891 in 2003-2008; 2074, 2827, 1803, 1682 compared to 3234 in 2009-2013), fewer breakdowns detected (158, 156, 151 and 150 compared to 161 in 2003-2008; 143, 143, 140, 140 compared to 145 in 2009-2013), higher number of latent infections (36, 38, 44 and 45 compared to 31 in 2003-2008; 34, 33, 38, 39 compared to 31 in 2009-2013), and a lower number of false positives (61, 81, 41 and 35 compared to 108 in 2003-2008; 45, 68, 41, 37 compared to 71 in 2009-2013). Finally, *Scenarios 7 and 8* detect a similar number of breakdowns as the *four-year* scenario (162 and 161 compared to 161 in 2003-2008; 146 and 145 compared to 145 in 2009-2013), have a similar number of latent infections (30 in both scenarios compared to 31 in 2003-2008; 30 in both scenarios compared to 31 in 2009-2013), present a higher number of false positives (136 and 120 compared to 111 in 2003-2008; 84 and 80 compared to 71 in 2009-2013), and test additional herds (5427 and 5173 compared to 4891 in 2003-2008; 3626 and 3486 compared to 3234 in 2009-2013).

Criteria	Points	Testing interval by points score	No. herds (%)	bTB (RHT)*
Better				
Slaughtering $< 25\%$ of stock	+1	0 points=no testing	2825 (24.7)	30 (2)
		1 point=4 year testing	8056 (70.4)	50 (10)
Receiving 'high risk' animals in > 3 years and slaughtering $< 50\%$ of stock	+1	2 points=2 year testing	556 (4.9)	17 (4)
Similar				
Slaughtering $< 25\%$ of stock and/or receiving 'high risk' animals in > 3 years and slaughtering $< 40\%$ of stock	1	0 points=no testing	2886 (25.2)	33 (2)
		1 point=4 year testing	8551 (25.2)	64 (14)
Lower Detection 1				
Slaughtering $< 12.5\%$ of stock and/or receiving 'high risk' animals in > 3 years and slaughtering $< 25\%$ of stock	1	0 points=no testing	4770 (41.7)	54 (5)
		1 point=4 year testing	6667 (58.3)	43 (11)
Lower Detection 2				
Slaughtering $> 25\%$ of stock	-1	-1 or 0 points=no testing	5064 (44.3)	19 (2)
Slaughtering $< 5\%$ of stock	+1	1 point=4 year testing	5241 (45.8)	58 (7)
Receiving 'high risk' animals in > 3 years	+1	2 points=2 year testing	1060 (9.3)	19 (7)
Having > 100 animals	+1	3 points = annual testing	72 (0.6)	1 (0)

Table 15: Composition of the risk-based surveillance scenarios for Scotland between 2003 and 2008. BTB, bovine Tuberculosis; RHT, Routine Herd Testing. BtB is the number of confirmed breakdowns between 2003 and 2008 that fell into that category. *RHT represents the breakdowns that were detected by RHT.

Criteria	Points	Testing interval	No. of herds (%)	bTB	RHT	WHT	SLH	Other
Scenario 1								
Slaughtering $< 25\%$ of stock	$+1$	0 points=no testing	4408 (22.5)	109	5	0	73	31
Receiving 'high risk' animals in > 3 years and slaughtering $< 50\%$ of stock	$+1$	1 point=4 year testing	13261 (67.8)	58	12	2	21	23
		2 points=2 year testing	1895 (9.7)	18	4	0	6	8
Scenario 2								
Slaughtering $< 25\%$ of stock and/or receiving 'high risk' animals in > 3 years and slaughtering $< 40\%$ of stock	1	0 points=no testing	4646 (23.7)	117	5	0	79	33
		1 point=4 year testing	14918 (76.3)	68	16	2	21	29
Scenario 3								
Slaughtering $< 12.5\%$ of stock and/or receiving 'high risk' animals in > 3 years and slaughtering $< 25\%$ of stock	1	0 points=no testing	8269 (42.3)	145	11	1	86	47
		1 point=4 year testing	11295 (57.7)	40	10	1	14	15
Scenario 4								
Slaughtering $> 25\%$ of stock	-1	-1 or 0 points=no testing	8805 (45)	43	5	1	17	20
Slaughtering $< 5\%$ of stock	$+1$	1 point=4 year testing	8724 (44.6)	126	11	1	77	37
Receiving 'high risk' animals in > 3 years	$+1$	2 points=2 year testing	1830 (9.4)	15	4	0	6	5
Having > 100 animals	$+1$	3 points=annual testing	205 (1)	1	1	0	0	0
Scenario 5								
Slaughtering $> 25\%$ of stock	-1	-1 or 0 points=no testing	13101 (67)	114	13	2	55	44
Slaughtering $< 5\%$ of stock	$+1$	1 point=4 year testing	5722 (29.2)	65	6	0	42	17
Receiving 'high risk' animals in > 3 years	$+1$	2 points=2 year testing	721 (3.7)	5	1	0	3	1
Having > 350 animals	$+1$	3 points=annual testing	20 (0.1)	1	1	0	0	0
Scenario 6								
Slaughtering $> 25\%$ of stock	-1	-1 or 0 points=no testing	13590 (69.5)	129	13	2	67	47
Slaughtering $< 5\%$ of stock	$+1$	1 point=4 year testing	5349 (27.3)	54	7	0	33	14
Receiving 'high risk' animals in > 3 years	$+1$	2 points=2 year testing	616 (3.1)	1	0	0	0	1
Having > 500 animals	$+1$	3 points=annual testing	9 (0)	1	1	0	0	0
Scenario 7								
≤ 350 animals	1	1 point=4 year testing	18849 (96.3)	160	17	2	83	58
> 350	3	3 points=annual testing	715 (3.7)	25	4	0	17	4
Scenario 8								
≤ 500 animals	1	1 point=4 year testing	19188 (98.1)	159	20	2	80	57
> 500	3	3 points=annual testing	376 (1.9)	26	1	0	20	5

Table 16: Composition of the risk-based surveillance scenarios for LRA England (2003-2008). The testing interval column represents the time frame of the bTB testing, which depends on the level of risk based on a score point system (0=no testing, 1=4 year testing, 2=2 year testing, 3=annual testing). The No. of herds column corresponds to the number of herds (and percentage of the total number of herds) that fell into each testing interval category. bTB is the number of confirmed breakdowns between 2003 and 2008. RHT, WHT, and SLH are the breakdowns that were detected by routine-herd testing, whole-herd testing and slaughterhouse, respectively.

Criteria	Points	Testing interval	No. of herds (%)	bTB	RHT	WHT	SLH	Other
Scenario 1								
Slaughtering $< 25\%$ of stock	+1	0 points=no testing	2632 (20.3)	93	1	1	62	29
Receiving 'high risk' animals in > 3 years and slaughtering $< 50\%$ of stock	+1	1 point=4 year testing	8356 (64.6)	34	4	0	12	18
		2 points=2 year testing	1951 (15.1)	19	4	1	5	9
Scenario 2								
Slaughtering $< 25\%$ of stock and/or receiving 'high risk' animals in > 3 years and slaughtering $< 40\%$ of stock	1	0 points=no testing	2815 (21.8)	94	1	1	62	30
		1 point=4 year testing	10124 (78.2)	52	8	1	17	26
Scenario 3								
Slaughtering $< 12.5\%$ of stock and/or receiving 'high risk' animals in > 3 years and slaughtering $< 25\%$ of stock	1	0 points=no testing	4642 (35.9)	106	2	1	68	35
		1 point=4 year testing	8297 (64.1)	40	7	1	11	21
Scenario 4								
Slaughtering $> 25\%$ of stock	-1	-1 or 0 points=no testing	4523 (35)	27	2	0	13	12
Slaughtering $< 5\%$ of stock	+1	1 point=4 year testing	6164 (47.6)	101	4	1	62	34
Receiving 'high risk' animals in > 3 years	+1	2 points=2 year testing	1932 (14.9)	14	3	1	4	6
Having > 100 animals	+1	3 points=annual testing	320 (2.5)	4	0	0	0	4
Scenario 5								
Slaughtering $> 25\%$ of stock	-1	-1 or 0 points=no testing	6827 (52.8)	72	3	0	38	31
Slaughtering $< 5\%$ of stock	+1	1 point=4 year testing	5074 (39.2)	67	6	2	41	18
Receiving 'high risk' animals in > 3 years	+1	2 points=2 year testing	1007 (7.8)	7	0	0	0	7
Having > 350 animals	+1	3 points=annual testing	31 (0.2)	0	0	0	0	0
Scenario 6								
Slaughtering $> 25\%$ of stock	-1	-1 or 0 points=no testing	7159 (55.3)	89	4	0	50	35
Slaughtering $< 5\%$ of stock	+1	1 point=4 year testing	4857 (37.5)	50	5	2	29	14
Receiving 'high risk' animals in > 3 years	+1	2 points=2 year testing	910 (7)	7	0	0	0	7
Having > 500 animals	+1	3 points=annual testing	13 (0.1)	0	0	0	0	0
Scenario 7								
≤ 350 animals	1	1 point=4 year testing	12417 (96)	124	8	2	65	49
> 350	3	3 points=annual testing	522 (4)	22	1	0	14	7
Scenario 8								
≤ 500 animals	1	1 point=4 year testing	12604 (97.4)	110	8	1	56	45
> 500	3	3 points=annual testing	335 (2.6)	36	1	1	23	11

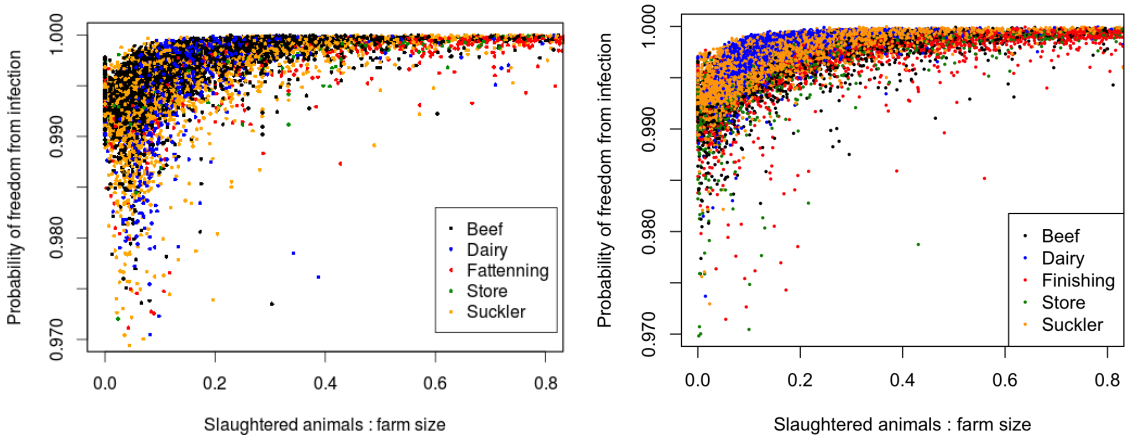
Table 17: Composition of the risk-based surveillance scenarios for LRA England (2009-2013). The testing interval column represents the time frame of the bTB testing, which depends on the level of risk based on a score point system (0=no testing, 1=4 year testing, 2=2 year testing, 3=annual testing). The No. of herds column corresponds to the number of herds (and percentage of the total number of herds) that fell into each testing interval category. bTB is the number of confirmed breakdowns between 2009 and 2013. RHT, WHT, and SLH are the breakdowns that were detected by routine-herd testing, whole-herd testing and slaughterhouse, respectively.

Tables 16 and 17 represent the composition of the risk-based surveillances where the testing interval is determined by the level of risk based on a score point system (0=no testing, 1=4 year testing, 2=2 year testing, 3=annual testing). However, *scenarios 7-8*, by definition, require the testing of additional herds, since they do not allow herds with high probability of freedom from the disease to be exempt from testing. In these scenarios, only herd size was accounted as a risk factor and all the herds below a certain threshold value (*scenario 7*: ≤ 350, *scenario 8*: ≤ 500) are tested every four years, while the ones above the threshold are tested every year. Tables 16 and 17 show that *scenarios 1-3* would have identified, respectively, 16, 16 and 10 out of 21 breakdowns that were identified by RHT for 2003-2008 and 8, 8 and 7 out of 9 for 2009-2013. While *scenarios 4-6* would have identified 16, 8 and 8, respectively, out of 21 breakdowns identified by RHT for 2003-2008, and 7, 6 and 5 out of 9 for 2009-2013. *Scenarios 7-9* would have identified all the breakdowns detected by RHT since all the herds are tested.

Factors for risk-based models

The characteristics of herds that have a low probability of freedom from slaughterhouse surveillance alone were analysed in order to identify the types of herd that require more surveillance. The following are analysed in relation to the probability of freedom from slaughterhouse surveillance only considering: 1) herd size; 2) herd type; 3) proportion of herds that is slaughtered; 3) receiving animals from HRAs either directly or indirectly.

When plotted against the probability of freedom of infection following slaughterhouse only surveillance, herds that slaughter a smaller proportion of their stock have a lower probability of freedom (Figure 16) for the different 3 scenarios. In LRA England, finishing herds (red dots) are the ones that send higher proportion of their stock to slaughter, while dairy (blue dots) and suckler (yellow dots) herds are the ones that send less. Dairy herds have a much higher probability of freedom from infection than any other type.



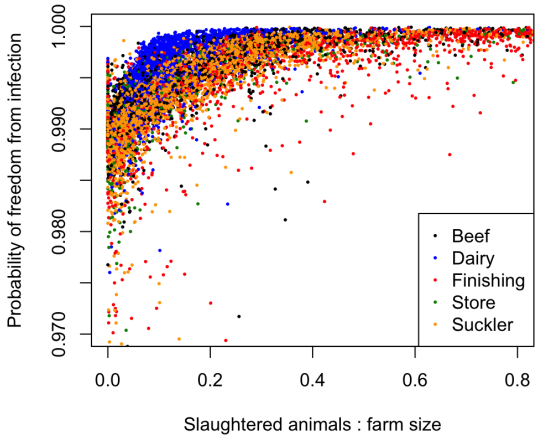


Figure 16: Plot of the ratio of slaughtered animals to probability of freedom from infection in Scotland (2003) (top left), in LRA England (2003), and in LRA (2009) (bottom). The axes on these plots have been truncated to ease interpretation.

To study which of the herd types have higher probability of freedom from infection, in this analysis, we will break down the probability of freedom by farm type according to SAM's database (Figure 17).

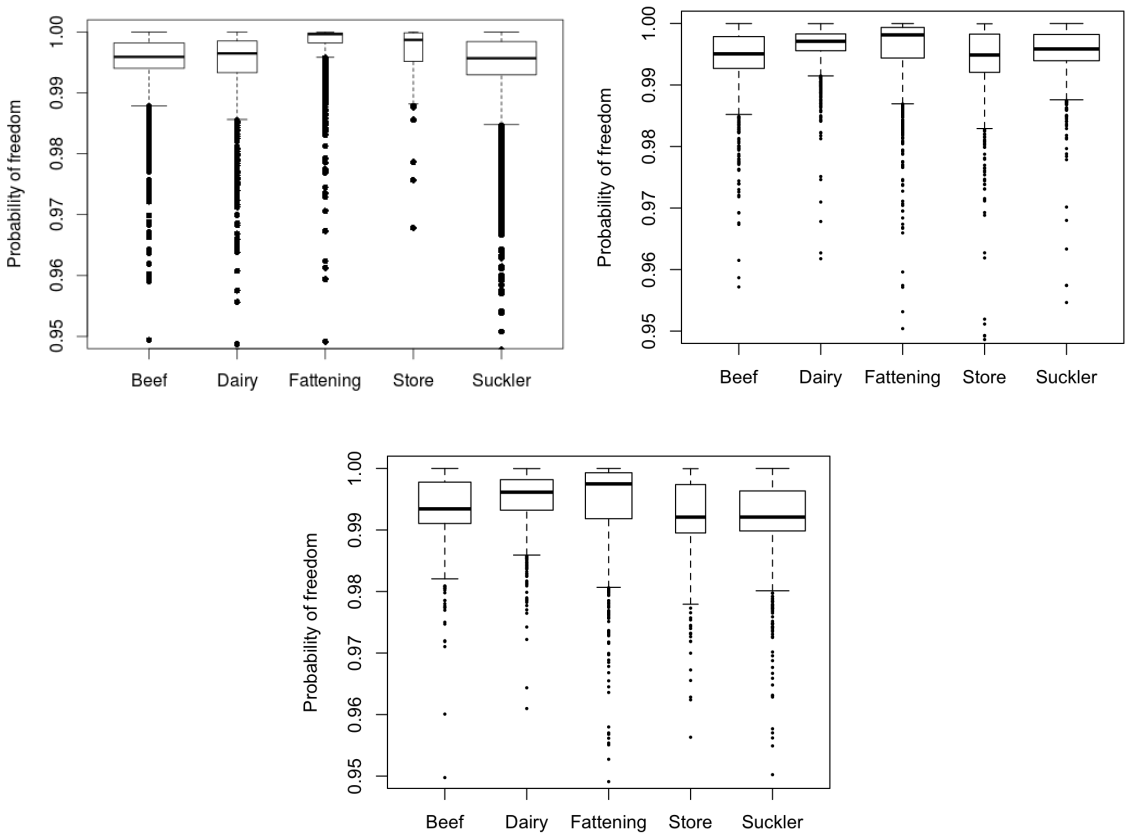


Figure 17: Boxplot of probability of freedom from the slaughterhouse only model for Scotland (top left), LRA England (top right) during 2002-2008, and LRA England during 2009-2013 (bottom) by herd type. Box widths represent the proportion of data in that category. The y-axis has been truncated for clarity.

Whilst not a particular risk group for acquiring infection, these figures show that the probability of freedom from infection is lower for suckler and store herds than for other herd types, for LRA England (2009-2013), while for Scotland and LRA England (2003-2008) the probability of freedom from infection is lower for suckler and store, respectively.

The width of the boxplots represent the number of herds that fit into each category. In Scotland (upper left), there are more beef herds (followed by suckler, dairy, fattening and stores), while in LRA England for 2003-2008, there are more suckler herds (followed by beef and dairy, fattening and lastly stores). In LRA England during 2009-2013, suckler herds continue to be the most abundant, however, beef, dairy and fattening herds are very close in number, followed by stores.

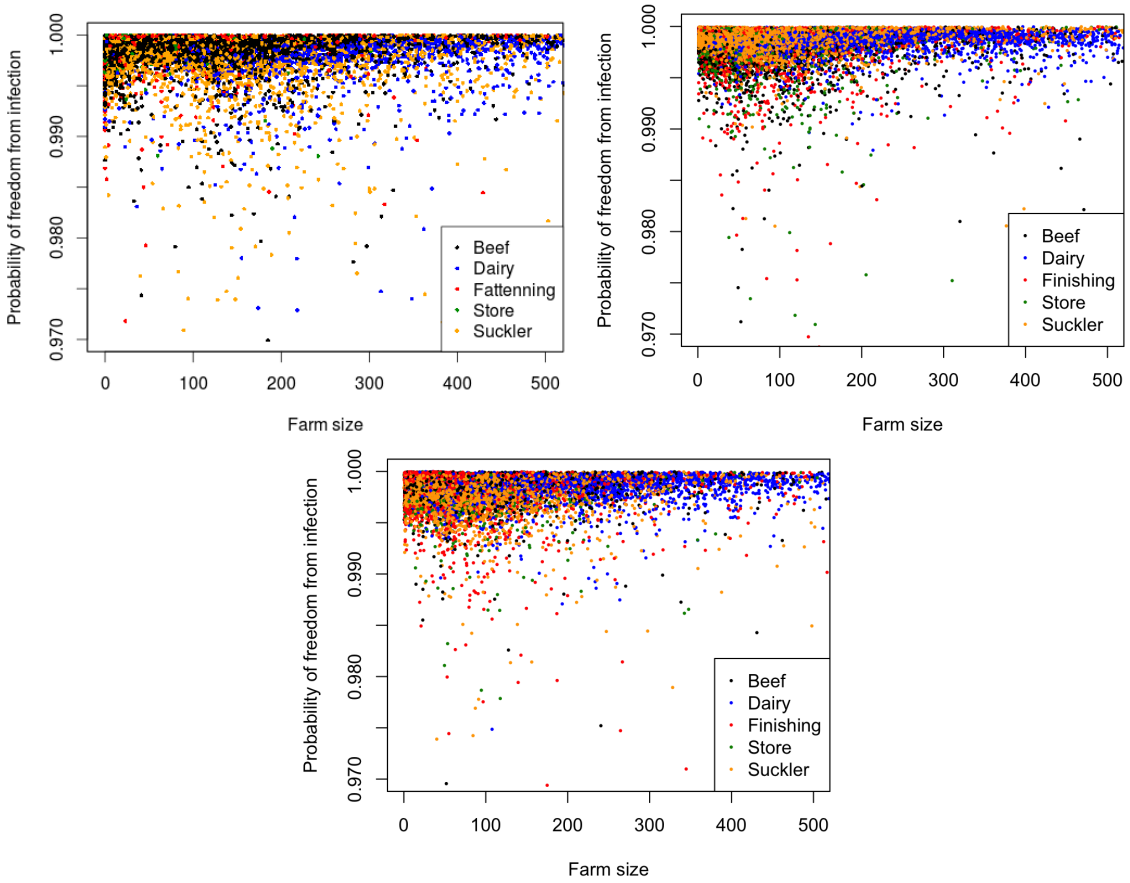


Figure 18: Scatterplot of the farm size categories by probability of freedom from slaughterhouse only surveillance in Scotland (top left), LRA England (top right) during 2003-2008, and LRA England (bottom) during 2009-2013. Axes have been truncated for clarity.

Herd size is an important risk factor, and Figure 18 shows that suckler herds are the ones that have smaller farm sizes and lower probability of freedom from infection (in orange), while dairy herds (in blue) are the ones that have larger farms and higher probabilities of freedom from infection in all three scenarios (less evident in Scotland).

An analysis of the number of years in which herds received animals from HRAs shows that the mean probability of freedom is insensitive to the number of years with high risk imports (see Figure 19). However, the tails of the distribution seem to be decreasing with the number of years with high risk imports. These results seem to be counter-intuitive since herds that receive animals from HRAs are at a higher risk, therefore more work needs to be done on the role of high-risk imports in cattle herds.

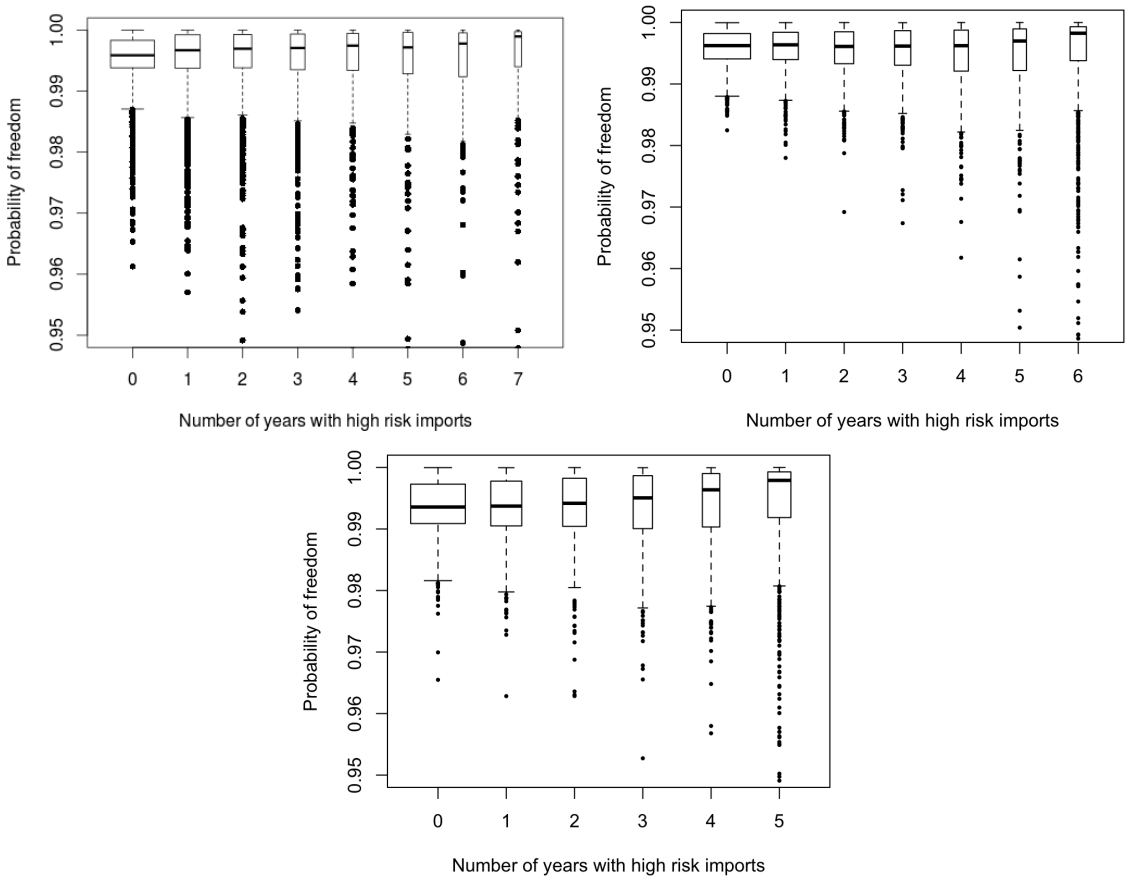


Figure 19: The probability of freedom by slaughterhouse surveillance in Scotland (top left) and in LRA England (top right) during 2003-2008, and LRA England (bottom) during 2009-2013, by the number of years in which animals were bought on to the farm from HRAs. The box widths represent the relative proportion of data points in each group. The y-axis has been truncated for clarity.

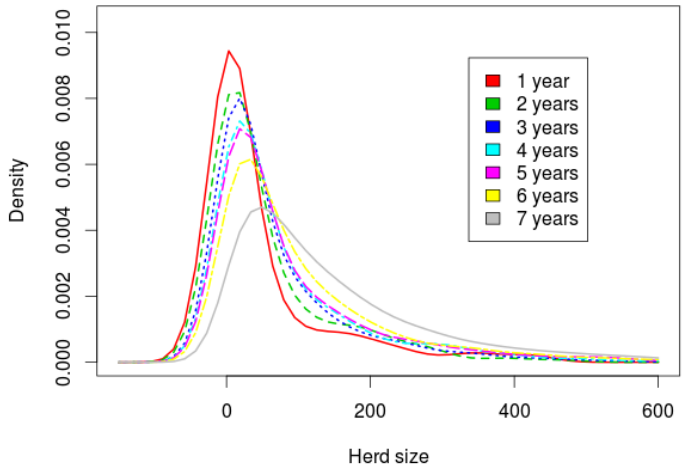


Figure 20: Herd size density of farms that imported animals from HRAs to Scotland. The number of years in which animals were bought on to farms from HRAs are separated by colour according to the legend.

To check how herd size is related with the number of years in which a farm has high risk imports, we compute the probability distribution of herd sizes per non-consecutive years of import. We expect that the longer a farm has been importing animals, the larger is the size of that farm. In Scotland, farms that import animals for longer periods are also farms that have more animals (Figure 20). By looking at the distribution of herd types per number of years that farms have been importing animals, Figure 21 shows that the proportion of herd types is broadly consistent with the number of importation years, except for beef fattening that increases its proportion when it has been receiving imports for 7 years.

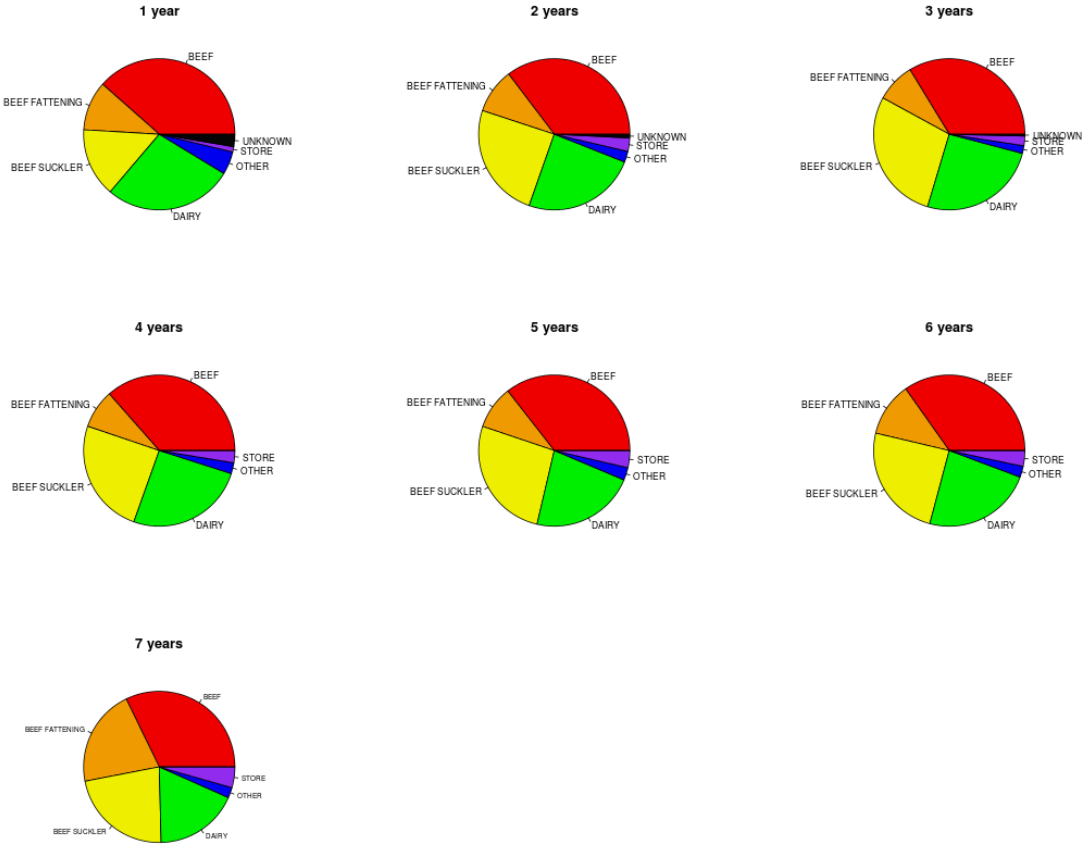


Figure 21: Herd type proportion of farms that imported animals from high incidence areas in Scotland. Each pie chart represent the number of years in which animals have been imported from high incidence areas. Herd types are represent by a different colour (beef:red, fattening:orange, suckler: yellow, dairy: green, store: violet, and other: purple).

This is further corroborated by the stump classifier models presented in this report, which also suggest that more precise targeting of movement types shows little benefit. Because volume of cattle movements is strongly correlated with herd size, it is likely that the importance of cattle movements is simply masked by this correlation. While the statistical analysis and, therefore, this result is used to establish the probability of freedom from infection, we emphasise the fundamental statistical principle that, while the correlation is useful for establishing relative risks, it does not imply causation. We therefore count cattle movements from high risk areas and from Ireland to be logical variables to consider in any risk analyses since they are likely to be the most important factor responsible for carrying infected individuals from herd to herd. Despite this decision, it does remain important to note the differences found for the risk factors in this study, compared to others. These differences could be due to the use of more conservative statistical methods, or

due to hidden relationships with other risk factors like herd size or herd type, masking in this way, their possible contribution.

We have applied the Scottish risk-based surveillance model developed by [12] to long term quadrennials in low risk areas in England.

In this model, four different baseline scenarios were modelled based on the annual time frame for surveillance (*slaughterhouse*, *four-year*, *two-year* and *one-year* models, where all have slaughterhouse surveillance together with the corresponding number of testing years). In the context of the Scottish framework, three risk-based surveillance scenarios were created for the cases that were identified as requiring fewer animals tests than the *current* surveillance regime (*better*, *similar* and *lower detection* surveillance) and evaluated relative to the number of latent infections.

This framework has been extended for LRAs England to incorporate one more baseline scenario (*two-year* testing) and four more alternative risk-based surveillance scenarios than the ones used in the Scottish framework. The strategies developed here provide different balances between the requirements for disease detection and minimizing surveillance effort, and the efficacy of detecting infections varies slightly when compared to the current *four-year* RHT. Eight surveillance scenarios were created to identify which testing regimes would require fewer animals tests, and fewer number of both latent infections and false positives in LRA England. The results show that *slaughterhouse* surveillance alone results in a large number of latent infections. Therefore, our study shows that additional active surveillance (via RHT) is necessary. A more frequent testing regime (one or two years interval) will not help in detecting more breakdowns than the current four year testing. These scenarios will decrease the mean number of latent infections, but they will increase substantially the number of false positives and of the number of herds tested.

Surveillance in LRA England can be improved if herds at higher risk are targeted. By exploiting statistical analyses defining herd level probability of freedom from disease stratified by significant risk factors, testing frequencies can be refined to test highly probable infections more frequently, and conversely potentially reducing testing of low risk herds. Of the scenarios developed, *scenarios 1-2* and *scenarios 7-8*, show very similar detection efficacy as the current testing regime and selected herds for testing based upon the proportion of stock slaughtered, the number of batches of high-risk animals moving into the farm (*scenarios 1-2*) and upon the herd size (*scenarios 7-8*). Scenarios 4-6 show lower detection efficacy than the current testing regime and included the testing of large herds (with different cut-off thresholds) as well as those that slaughter fewer animals and import animals from high-incidence areas. Different temporal windows were investigated in this analysis: one-, two-, four-year testing, and combinations of the formers. Different solutions emerged from these testing windows. The results show that if greater detection to achieve freedom than the current regime were desirable, then testing herds at a greater risk of acquiring infection on a more regular basis may be beneficial, but at some cost in terms of numbers of herds being tested and an increase in the number of false positives. As determinants of risk and subsequent detection we have used herd size, proportion of animals sent to slaughter and movements from high risk areas in England and Wales, and from both Republic of Ireland and Northern Ireland. Two of the systems presented here require testing herds that slaughter $< 25\%$ of stock per year. This ensures that those herds that would not on average replace their entire stock over a 4-year period are tested for bTB. Five of the systems require testing herds that have large numbers of cattle. Herd size has been a known risk factor for bTB and a more frequent testing regime in these herds would improve detection of disease.

During the periods 2003-2008 and 2009-2013 there were 185 and 146 breakdowns, where 21 and 9 of the cases were identified by RHT, 101 and 79 by slaughterhouse, 2 and 2 by WHT. Despite comprising only RHT and slaughterhouse surveillance, the model predicts around 161 and 145 of cases detected for the periods 2003-2008 and 2009-2013, respectively. This is because the remainder

of the surveillance is made up of other types of testing not made up for here, such as pre- (not mandatory for LRA in England) and post-movement tests, post-Irish imports, tracings, hotspots check and contiguous herd tests. These are not included in the model, however, it still allows for some of their detection indirectly at slaughterhouse or by RHT of those that would occur later in time. Of the eight scenarios developed for LRA England, there were three scenarios that produced similar levels of surveillance (the mean number of latently infected herds is within 5% of the current surveillance) for fewer herds tested (Scenario 1) and for additional herds tested (*scenarios 7-8*), and five scenarios that produced up to 15% lower surveillance (*scenarios 2-6*). Despite this, they can offer different advantages. For 2009-2013, *Scenarios 2-3*, would miss 1 and 2 breakdowns that were detected by RHT (by not testing all the herds), but would involve fewer tests than *scenarios 4-6*. *Scenario 4* would miss 2 breakdowns while *scenarios 5-6* would miss 3 and 4, respectively that were detected by RHT. However, *scenarios 5-6* would test fewer herds than *scenario 4* (*scenario 6* would test even fewer than *scenario 5*). Finally *scenarios 7-8*, would not exempt any herds for being tested, 'penalising', this way, large herds with a more frequent testing regime, which could serve as a disincentive to this behaviour. These scenarios provide a better surveillance than the *scenarios 2-5*. Both scenarios would identify the 9 breakdowns detected by RHT, and *scenario 8* would test fewer herds. However, by comparing the scenarios that had similar levels of surveillance to the current one, *scenario 1* would miss 2 breakdowns that were detected by RHT, while *scenarios 7-8* would not miss any, but at the cost of testing more herds. In summary, *Scenario 1* and *2* would keep the same number of detections and latent infections, would decrease substantially the number of false positives, and would offer savings by testing fewer herds. *Scenarios 7-8* would also detect a similar number of breakdowns and number of latent infections, however, they would increase the number of false positives, and test more herds than the current system. This model accounts only for slaughterhouse surveillance and RHT, and to guarantee that extra breakdowns will be detected, other types of tests that have been currently in place must be continued to be implemented.

4. **Use mathematical models to determine consequences of possible missed breakdowns.**

   A Bayesian likelihood model developed by O'Hare and Colleagues ([20]) has been used to determine the consequences of reduced testing, and to identify possible epidemiological differences in the transmission of bTB across herd types.

   Figure 22 shows the distributions of the breakdown sizes in LRAs used to estimate the parameters for each herd type (beef and dairy) used in the within-herd model. Breakdown sizes are defined as the number of reactors at first breakdown. Only breakdowns that were detected by RHT or tests triggered by slaughterhouse detection were included. Furthermore, there were only included herds that did not have a previous breakdown. Analysis show that smaller breakdowns are more frequent than larger breakdowns, with the number of beef herds that experience a break down almost double the number of dairy at lower herd size, but with the discrepancy reduced when one compares herds of larger extent.

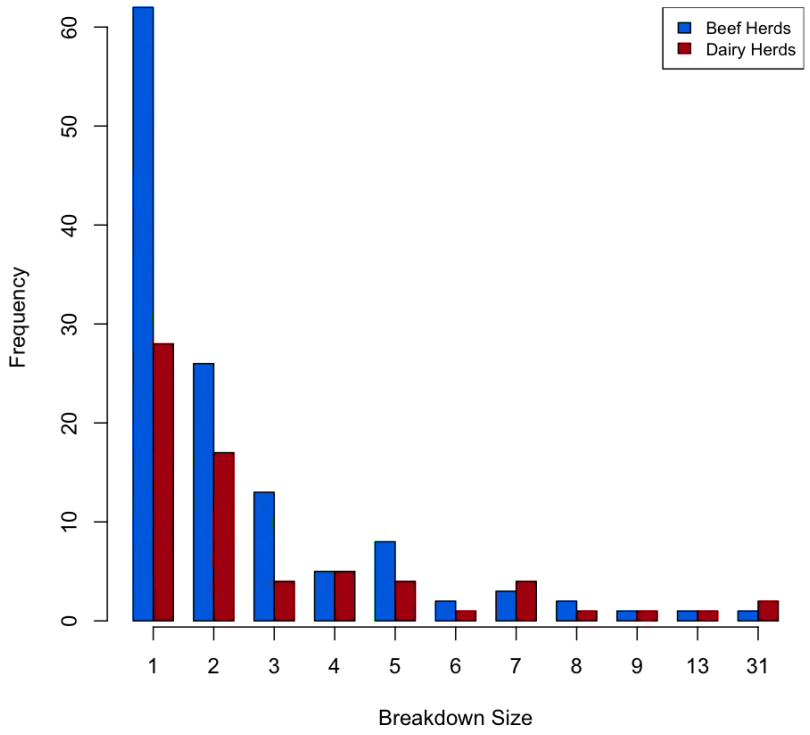


Figure 22: Frequency distributions of breakdown size for beef (blue) and dairy (red) herds in 2006. The breakdown size is defined as the number of reactors at first breakdown. Only breakdowns that were detected by RHT or tests triggered by slaughterhouse detection were included.

Derivation of the distributions used in the model

The age distribution used in the model was derived from the distribution of animal ages at the time of a positive or negative test in GB in 2006 (as recorded in VetNet) and separated by herd type (where the herd type is recorded as 'DAIRY' or 'BEEF' in VetNet) and the distribution of ages of animals presented to slaughter (Figures 23 and 24). We identified 3 age groups from these distribution and are given in Table 24.

From the age distribution of the national herd in 2006 we calculate the fraction of animals in each identified age group. In each simulation we select a herd size and calculate the numbers of animals in each age group from this.

We calculate the distribution of herd sizes from the size of each herd when it first suffered a breakdown (i.e. using the same criteria used to obtain the breakdown size distribution to define our likelihood) but also including herds that tested clear during the same time period. We did not include follow-up tests due to inconclusive reactors.

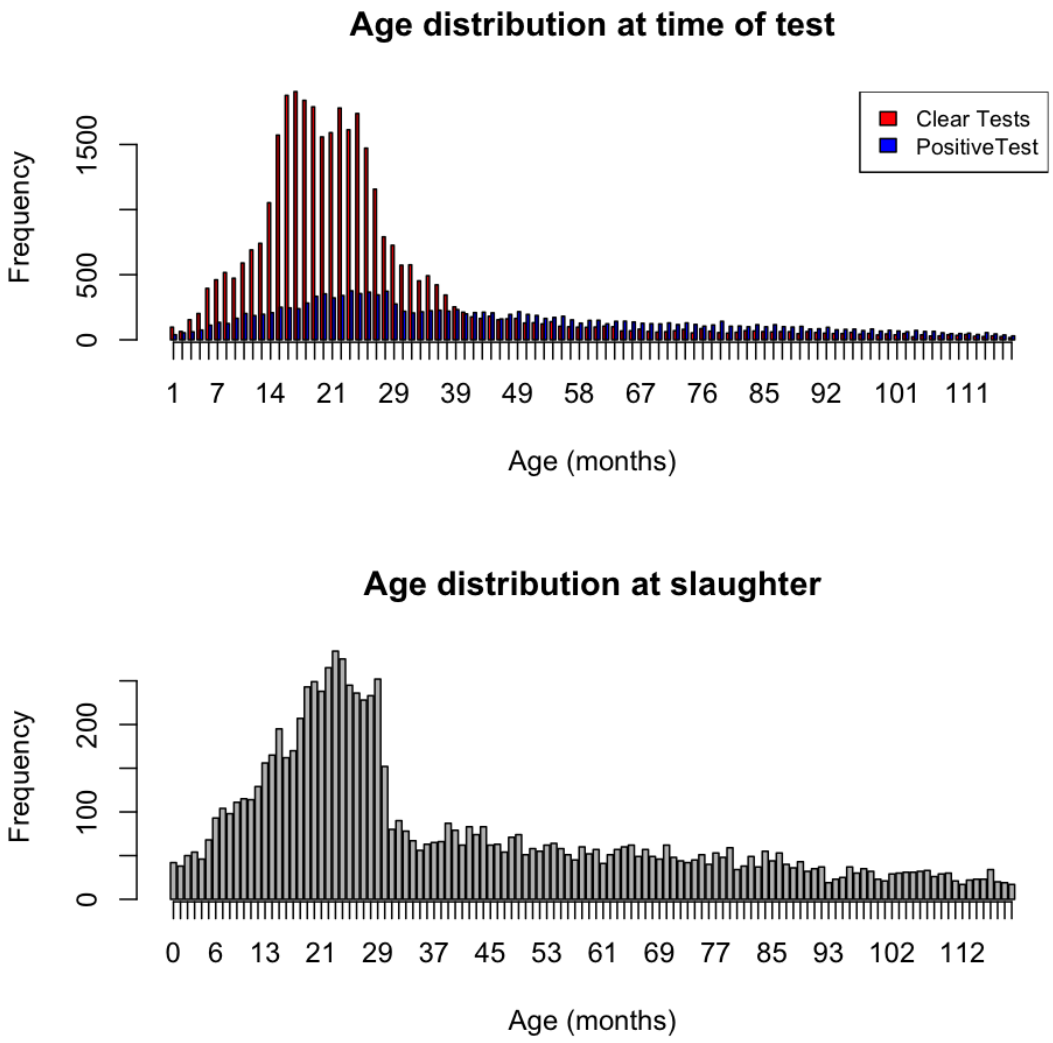


Figure 23: Age distributions of beef herds in GB in 2006 used to determine the age structure for the model. The age of of the animals at the time of test (top) is calculated as the proportion of all tests that either test positive or clear, inconclusives are ignored. The age distribution at slaughter (bottom) is obtained from the age of the animals when presented at slaughter.

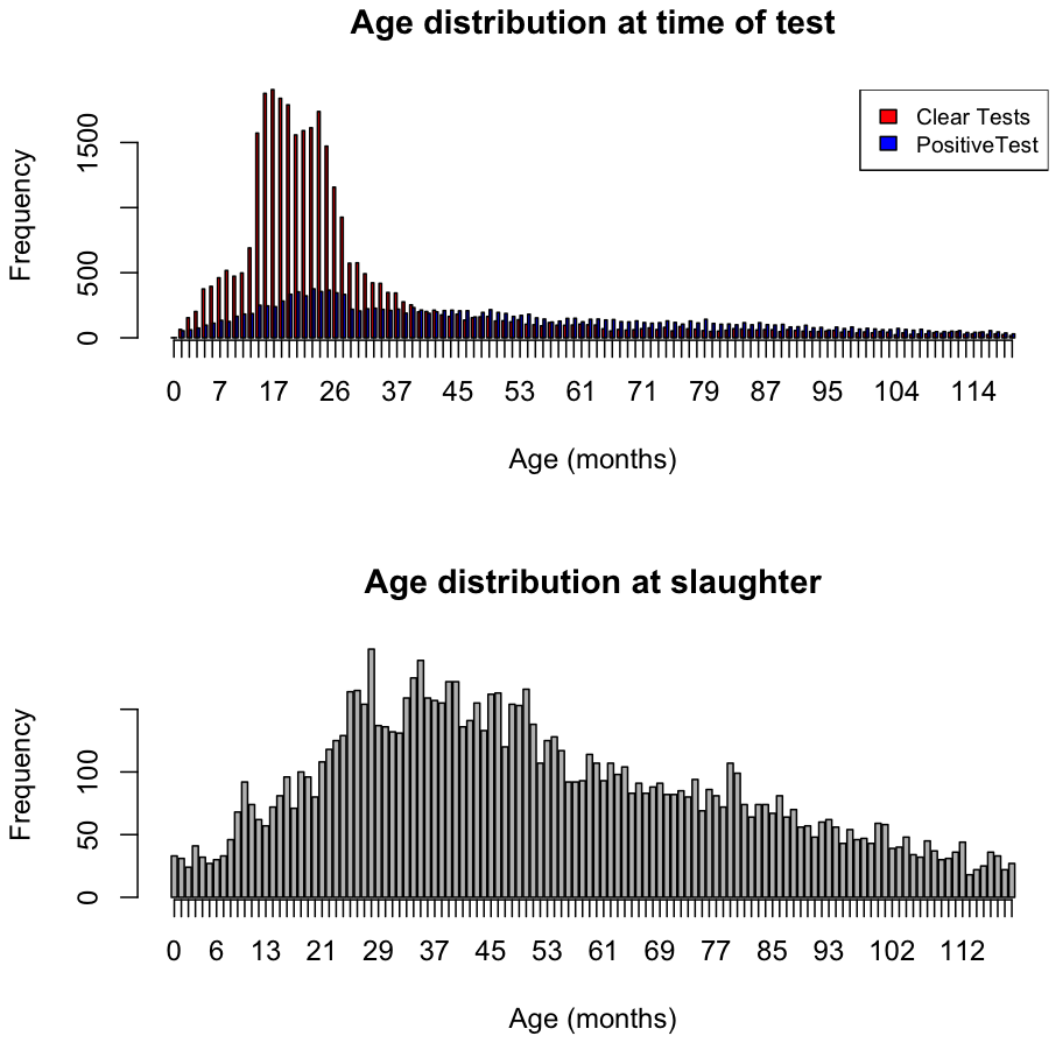


Figure 24: Age distributions of dairy herds in GB in 2006 used to determine the age structure for the model. The age of of the animals at the time of test (top) is calculated as the proportions of all tests that either test positive or clear, inconclusives are ignored. The age distribution at slaughter (bottom) is obtained from the age of the animals when presented at slaughter.

Age (months)	%age of Dairy Herds	%age of Beef Herds
0 - 14	24.7%	49.5%
15 - 28	25.7%	39.4%
29 - 39	22.7%	4.1%
> 40	26.7%	7.2%

Table 18: **Age distribution as a percentage of the overall herd size.** The age distribution of the herd is determined by considering the age distribution of the GB herd as a whole, the age distribution of animals presented for routine testing and the age distribution of animals at slaughter.

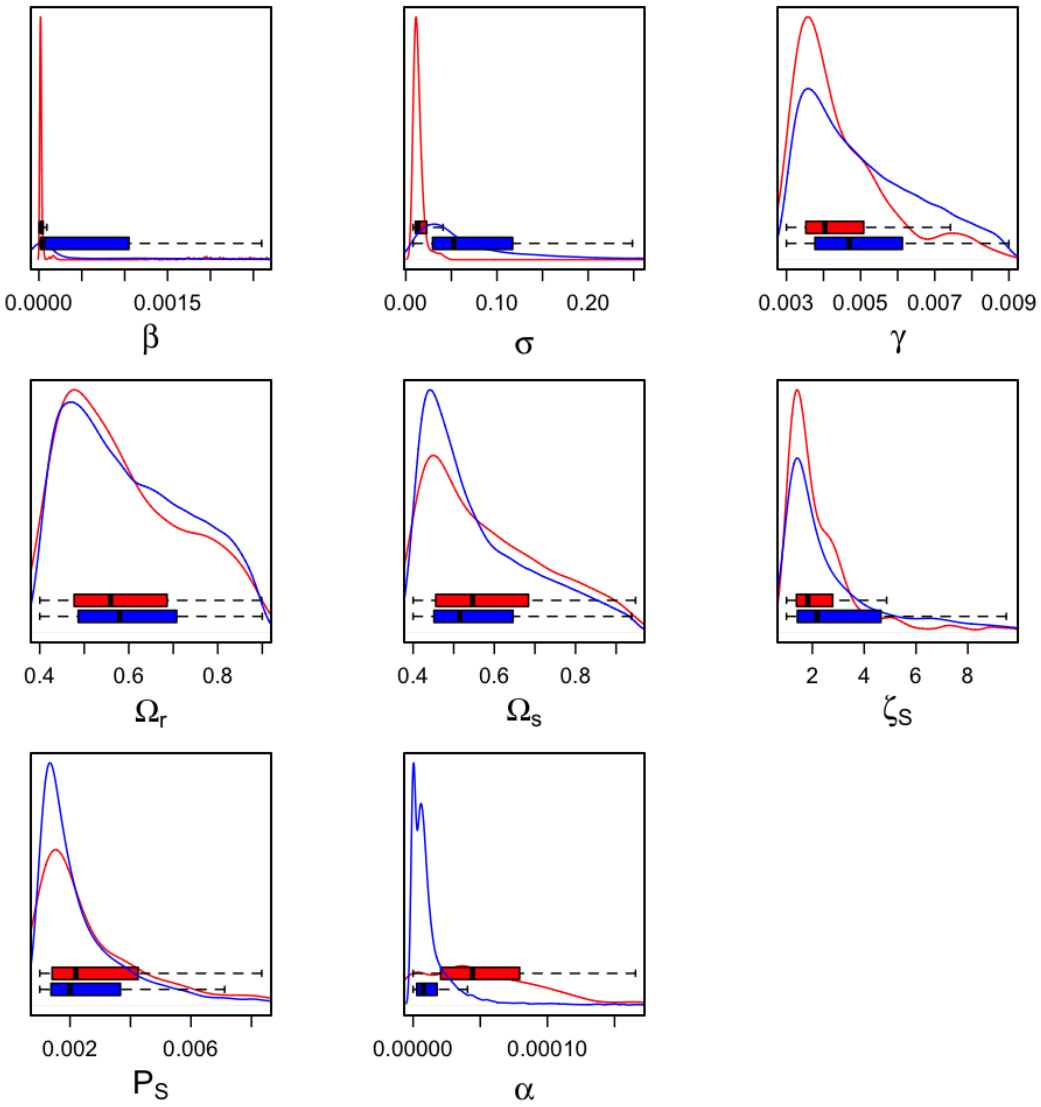


Figure 25: Posterior distribution of the sampled parameters. Parameter description: β: transmission rate; σ: rate of exposed cattle becoming test sensitive; γ: rate of test sensitive cattle becoming infectious; Ω_r: probability that a test sensitive or infectious; Ω_s: probability that a test sensitive or infectious animal is detected at abattoir; ζ_S: increased infectiousness of superspreaders; P_S: proportion of individuals that are superspreaders; ζ_H: increase of β in herds with high β; P_H: proportion of herds with high β; and α: external force of infection. Beef herds are represented in blue and dairy herds in red.

Figure 25 represents the posterior distribution of all the parameters sampled for both beef (blue colour) and dairy (red colour). The distributions of each parameter suggest that dairy herds are more susceptible to an external force of infection (subplot related to parameter α) and that beef herds might have a higher rate of exposed cattle becoming test sensitive (subplot related to parameter σ). All the other parameters seem to have a similar pattern when compared between beef and dairy herds.

Figures 26, 27 and 28 show results of simulations where RHT have been used and where there is slaughterhouse surveillance only. These simulations were run only on herds that were persistently in LRAs for the entire study period and where there was not a previous breakdown in the herd. Here, both beef and dairy herds were analysed together. By having slaughterhouse alone as a testing regime, the breakdown size would be bigger than if RHT had been implemented (Figure 26).

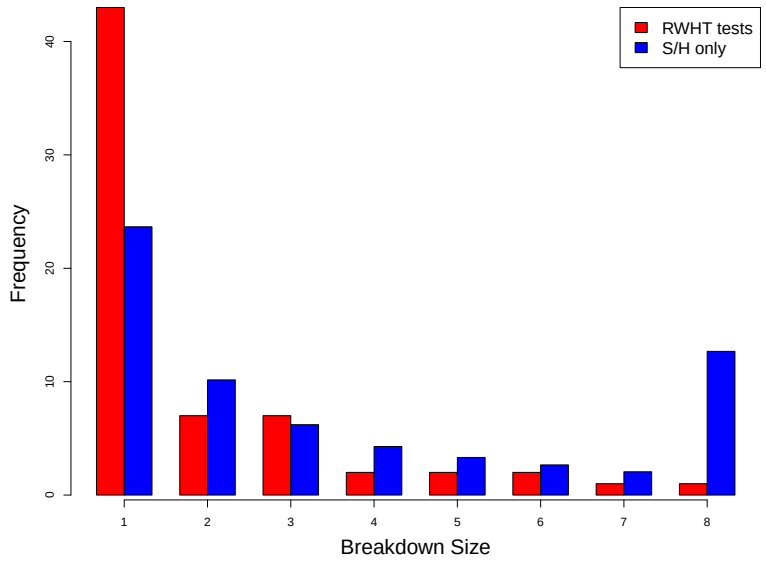


Figure 26: Frequency distribution of the breakdown size (the number of reactors at first breakdown) for dairy and beef herds together in persistent LRAs with RHT (red colour) and in the absence of RHT (slaughterhouse only)(blue colour). The distributions are the result of simulations using data from dairy and beef herds in GB.The number 8 on the x-axis represents ≥ 8 reactors.

Breakdown duration (episode length) with slaughterhouse alone would be longer than if RHT were implemented, where 100% restrictions would have lasted 6 months with RHT, while they would have lasted 20 months without it (Figure 27).

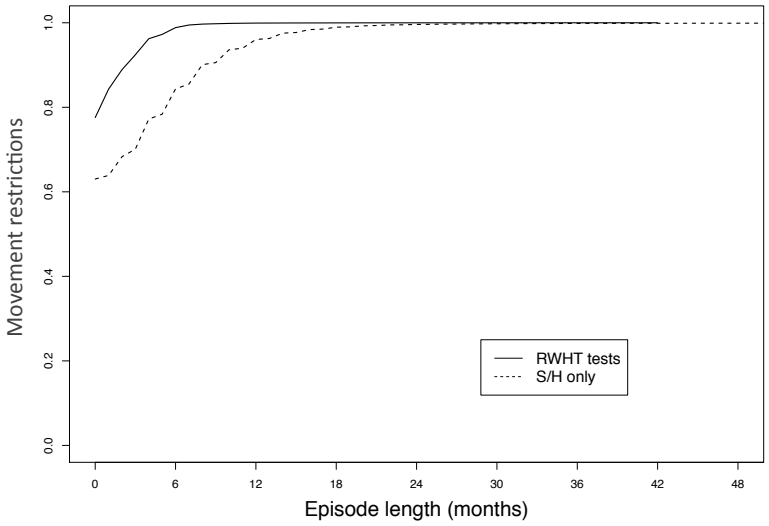


Figure 27: Breakdown length with RHT (black line) and with slaughterhouse testing only (dashed line) versus movement restrictions. The x-axis represents the episode length in months, while the y-axis represents the percentage of movement restrictions that have been imposed.

Slaughterhouse alone would take much longer to detect infection than RHT and it would detect less than 50% of cases (Figure 28).

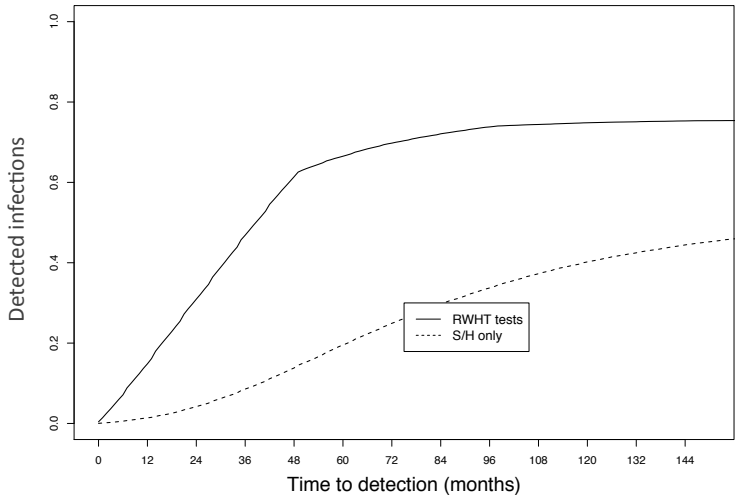


Figure 28: Breakdown length with RHT (black line) and with slaughterhouse testing only (dashed line) versus detected infections. The x-axis represents the episode length in months, while the y-axis represents the percentage of detected infections.

Discussion

This report has provided recommendations for improved surveillance for bTB that was presented in four separate stages:

Identification of areas that could be epidemiologically classified as HRA, LRA and TA.

In this section, we used the predictive model developed by [13] to exploit the spatial clustering of *M.bovis* genotypes and gain insight into which geographical regions British cattle are at high risk of becoming infected. This estimate was based on using a combination of spatial, network and genotype information to explain the patterns of outbreaks and identifying when cattle movements are insufficient to explain outbreak patterns, therefore requiring they be explained by spatially local risks. Based on results from a previous Defra project (SE3243), the optimised HRA detection algorithm provides a more robust way of identification of HRAs and therefore by extension, of transitional areas. Based on this new definition of high risk quadrats, we were able to detect geographical infection before their perceived risk (identified via implementation of annual testing), and to more rigorously identify future transition areas. We have classified transitional areas as any that became high risk or low risk (after already being considered high risk) during the study period. The disparity between the historical testing regime and our attributions of HRAs suggest that the latter may be useful for refining future epidemiological analyses, and establishing the best means of controlling the spread of bTB.

Identification of statistical risk factors

As the incidence of bTB continues to rise, it is important to develop an improved understanding of the characteristics of the herds that are becoming infected in areas of different levels of risk. This can aid control and surveillance of the disease to help to ensure that the disease is contained with the more suitable surveillance strategies possible. The analyses presented in this report identify herds that are more likely to have a bTB breakdown. Variables associated with the risk of infection appeared to depend upon the local of incidence of the disease. The following risk factors were considered for analysis: herd type, herd size, movements from HRAs in England and Wales, Irish imports, and predicted imputed badger density. For both LRAs and HRAs, dairy herds have a similar risk profile to beef, however, for the earlier period of LRAs they do not seem to be statistical significant. Finishing herds in HRAs are the ones that have higher risk profile. Herd size was presented as a risk in all areas, but much higher in LRAs suggesting that there might be a relationship between herd size and cattle imports from HRAs, as bought in cattle are the most likely underlying associated risk factor, based both on our own results and previous investigations. Imputed badger density only showed up as risk factor in LRAs and for the later study period, which suggests that there might related with other local processes. Movements from HRAs were considered as being a significant risk factor for all the areas. While they present a high risk for LRAs, in HRAs they seem to have no relationship with an increase of risk since the herds are already at a high risk of infection. In TAs, it seems that there is also not an increase of risk associated with cattle movements from HRAs, which will need further exploration. Imports from Ireland only showed up as a risk factor in the later period of LRAs, suggesting that their impact might be masked by other factors in the earlier period. Any attribution of the raw risk factors epidemiologically (rather than as an identification of correlation to likelihood of becoming infected) should therefore be used with caution.

Identifying outcomes of revised surveillance strategies

By adopting a previously developed surveillance model that was used to identify targeted strategies in Scotland, for herds in LRAs in England, a variety of testing regimes were considered with the aim of identifying the impact on the breakdown identification under different surveillance levels. Our analyses show that slaughterhouse surveillance leaves many unidentified, infected herds, and in contrast testing all herds more frequently (every year or every two years) will also not increase the chances of detecting many more infections due to the sensitivity of the skin test. It would also substantially increase the cost of testing and would increase the number of false positives, since many more herds would be tested, however, targeting some herds come be beneficial. Some of the risk-based surveillance scenarios seem to offer advantages over others. *Scenarios 1* and *2*, which penalize herds that send a small proportion of their stock to slaughter and that receive animals from high risk areas for more than three years, seem to reduce the number of herds tested, the number of latent infections and the number of false positives for a similar level of detection as the current four-year surveillance regime. Other also possible beneficial scenarios would be *scenarios 7* and *8*, which penalize large herds and for a very similar level of detection as the current regime can reduce the number of latent infections. Because these scenarios test all herds either every year (very large herds) or every four years (smaller herds), the number of tested herds is substantially higher than the current regime, increasing the number of false positives and, subsequently, the cost associated to the implementation of this system. Even if the number of detected infections does not improve by applying one of these systems, they are still very relevant not only because of cost-related benefits but because they would reduce the number of latent infections, which would avoid the hidden spread of the disease and make a huge impact on the future control of bTB in LRAs. While there are potential savings to be made, we would stress the need to regularly monitor the fit of outbreak distributions to risk factors identified here and elsewhere. By their nature, the impact of correlation without necessarily causation means that the statistical relationship between identified risks and outcomes can change or be broken down entirely, even when the identified risk factors are also epidemiologically sensible. These changes can be due to many factors, most importantly human behaviour, where individuals may respond to changes in regulations and social factors in unexpected ways.

Within-herd models of bTB: Determine consequences of possible missed breakdowns and identifying epidemiological risk factors for bTB infected herds

We used within-herd models of bTB transmission for two purposes first, to determine the consequences of reduced testing, and second, to identify possible epidemiological differences in the transmission of bTB across herd types. One of the consequences of a reduction in levels of active surveillance, is the possibility of delays to detection of disease. Dynamic models of disease transmission are one way of estimating the duration of such delays, and our models suggest that delays may be considerable, with undetected infections lasting on average approximately twice as long as in the absence of RHT. This is not likely to be an issue where RHT-exempt herds sell on relatively few cattle to other herds, and may be mitigated by additional measures, for example, requiring pre-movement testing. In addition, the increased number of infected cattle that are likely with such delays are also important, due to their impact on farmer livelihood, however, the model suggests that while there are delays, even without RHT, most herds return to OTF status quickly. Fitting these dynamics models (in this case, in HRAs) also suggests that there is little to distinguish the within-herd transmission characteristics of beef and cattle herds and, therefore, little evidence on which to base targeting surveillance on the basis of onward risk to other herds. While these clearly experience different patterns of outbreak sizes when testing positive, the model results are consistent with these differences being due to other things than transmission, in particular, it appears that the dominant factor is the risk of exposure to bTB from sources outside the herd. It is notable that the result that dairy herds are more exposed to external risk than beef herds is also consistent with our statistical result that dairy herds experiencing breakdowns more frequently is

an indicator of an incipient HRA, and may be worthy of further investigation.

This report provides a series of recommendations based on different bespoke analyses to determine the value of risk-based surveillance where herds are at high risk and low risk of bTB infection, and areas that are transitional between the two. The recommendations are largely based on combined mathematical and statistical models fitted to the detailed outbreak data that are available in GB. A critical consideration when reviewing these recommendations is the knowledge that the landscape of bTB transmission to cattle herds in GB is always changing, because of natural factors (the epidemiology of bTB in cattle and badgers), changes in the industry, and due to responses to changes in legislation. As such, while in our view the principles behind the recommendations are sound and (as they are consistent with epidemiological information) are likely to remain robust, the specifics of the recommendations would require regular review and re-analysis.

Conclusion

The key outcomes of our analyses are:

1. For herds in LRAs in England, movements from HRAs are a better predictor of risk of a breakdown than for herds in Scotland (where a movement-based risk-surveillance strategy is already in place). Further statistical analysis suggests that other factors such as herd size may be better predictors of risk but these are highly correlated to movement volumes (and so this outcome is likely the result of statistical correlation). As cattle movements are the most likely cause of this risk, further work on targeting surveillance is based on a movement-based risk model. More intensive testing of some high risk cattle herds appears to have little advantage, however a targeted surveillance would still be worthwhile with a reduction or an increase of the number of herds tested. Therefore, it would be advantageous to develop a cost-benefit analysis to validate these results.

2. For herds in HRAs in England, there is increased outbreak risk in finishing herds. There is also suggestion that large herds are at a higher risk, however, there is not much difference between the last two herd category sizes.

3. For herds in TAs in England and Wales, herd size and movements from high risk areas were suggested as the most significative risk factors. Generally, herds of the size $351 - 500$ seem to be the ones more at risk.

Our analyses suggest that there are useful ways in which surveillance can be targeted across all three types of areas (LRAs, TAs and HRAs). However, some caveats remain:

1. Any identification of targeted strategies is more robust if the targeting is at the epidemiologically significant factors, rather than correlated factors. As these models are (largely) statistical only, external evidence on their epidemiological importance must be considered when choosing the most likely candidates for targeting factors, rather than simply best statistical fit.

2. In particular for TAs, because the epidemiological factors for this area are poorly understood and may differ across regions, any results must be treated with particular caution. As well, the relatively small size of the areas considered and few herds means that the statistical significance is harder to identify.

3. Any change in surveillance may have unintended consequences for farmer behaviour and these must also be carefully considered.

4. While imputed badger density was a considered risk factor, the counterintuitive outcomes, and possible issues associated with interpreting these outcomes have led us to reject further use of them in the risk factor model, and more generally suggest some caution should be extended to using these density estimates as risk factors for bTB without further validation.

5. The Scottish-risk based surveillance model presented different results from the ones known to the Scottish data for the period 2003-2008, mainly related to the *better* (renamed as *scenario 1* for the LRA England study) surveillance system, which showed similar number of detected infections than the current surveillance scenario. This suggests that the epidemiological signatures between Scotland and LRA England might be very different. In particular, because the factors that are most correlated with high breakdown risk in LRAs need further investigation to establish their epidemiological significance (most importantly, the role of imputed badger density), more research of the underlying risk factors must be considered before establishing more targeted and or more

intensive surveillance schemes. As the framework used is scalable, has a flexible configuration and is adaptable, it is straightforward to modify or incorporate other surveillance scenarios, as well as possible future policy developments, such as the introduction of post movement testing, for example.

Further work:

- Identify TAs for all spoligotypes including recent years to improve statistical power.
- Epidemiological study to consider dairy herds as potential sentinels for incipient high risk areas.
- Investigate relationships between risk factors to determine the imputed badger density models, and cattle herd bTB risk factors.

References

[1] Kao RR, Gravenor MB, Charleston B, Hope JC, Martin M, et al. (2007) *Mycobacterium bovis* shedding patterns from experimentally infected calves and the effect of concurrent infection with bovine viral diarrhoea virus. 4: 545–51.

[2] Gopal R, Goodchild A, Hewinson G, de la Rua Domenech R, Clifton-Hadley R (2006) Introduction of bovine Tuberculosis to north-east England by bought-in cattle. Veterinary Record 159: 265–271.

[3] Defra: TB in cattle http://www.defra.gov.uk/statistics/foodfarm/landuselivestock/cattletb/national/

[4] Brooks-Pollock E, Keeling M (2009) Herd size and bovine Tuberculosis persistence in cattle farms in Great Britain. Preventive veterinary medicine 92: 360–5.

[5] Conlan AJK, McKinley TJ, Karolemeas K, Brooks-Pollock E, Goodchild AV, et al. (2012) Estimating the Hidden Burden of Bovine Tuberculosis in Great Britain. PLoS Computational Biology 8: e1002730.

[6] Gillespie D (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. Journal of Computational Physics 22: 403–434.

[7] Diekmann O, Heesterbeek JAP, Metz JAJ (1990) On the definition and the computation of the basic reproduction ratio r_0 in models for infectious diseases in heterogeneous populations. Journal of Mathematical Biology 28: 365-382.

[8] Hastings WK (1970) Monte Carlo Sampling Methods Using Markov Chains and Their Applications. Biometrika 57: 97.

[9] Clegg TA, Good M, Duignan A, Doyle R, Blake M, et al. (2011) Longer-term risk of *Mycobacterium bovis* in Irish cattle following an inconclusive diagnosis to the single intradermal comparative tuberculin test. Preventive Veterinary Medicine 100: 147–54.

[10] Gates MC, Volkova VV, Woolhouse MEJ (2013) Risk factors for bovine tuberculosis in low incidence regions related to the movements of cattle BMC Veterinary Research 2013, **9**:225 doi:10.1186/1746-6148-9-225

[11] Downs SH, Parry J, Nunez-Garcia J, Abernethy DA, Broughan JM, et al. (2011) Meta-analysis of diagnostic test performance and modelling of testing strategies for control of bovine Tuberculosis in GB. In: Proceedings of Society for Veterinary Epidemiology and Preventive Medicine.

[12] Bessell PR, Orton RJ, O'Hare A, Mellor DJ, Logue D, et al. (2012) Developing a framework for risk-based surveillance of Tuberculosis in cattle: a case study of its application in Scotland. Epidemiology and infection : 1–10.

[13] Orton RJ, Bessell PR, Green DR, Kao RR (In prep.)Using genotype and explicit contact data to identify high-risk areas for bovine Tuberculosis in British Cattle.

[14] Etherington TR, Ward AI, Smith GC, Pietravalle S, Wilson GJ (2009) Using the Mahalanobis distance statistic with unplanned presence-only survey data for biogeographical models of species distribution and abundance: a case study of badger setts Journal of Biogeography: 36, 845853.

[15] Gilbert M, Mitchell A, Bourn D, Mawdsley J, Clifton-Hadley R, Wint W (2005) Cattle movements and bovine tuberculosis in Great Britain. Nature 435: 491–496.

[16] Green DM, Kiss IZ, Mitchell AP, Kao RR (2008) Estimates for local and movement-based transmission of bovine Tuberculosis in British cattle. Proceedings of the Royal Society B 275: 1001–1005.

[17] Bates D, Maechler M, Bolker B and Walker S (2014) lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7 http://cran.r-project.org/web/packages/lme4.

[18] Tremblay A, and Ransijn J, (2012) LMERConvenienceFunctions: a suite of functions to back-fit fixed effects and forward-fit random effects, as well as other miscellaneous functions. R package version 2.5 http://cran.r-project.org/web/packages/LMERConvenienceFunctions.

[19] Bessell P, et al. (2012) Risk factors for bovine Tuberculosis at the national level in Great Britain BMC veterinary research 8.1: 51.

[20] O'Hare, A., R. J. Orton, P. R. Bessell, and R. R. Kao (2014) Estimating epidemiological parameters for bovine tuberculosis in British cattle using a Bayesian partial-likelihood approach Proceedings of the Royal Society B: Biological Sciences: 281, no. 1783

[21] Downs SH, et al (2011) Meta-analysis of diagnostic test performance and modelling of testing strategies for control of bovine tuberculosis in GB Proceedings of SVEPM. Leipzig, Germany: 139-153

[22] Judge J, Wilson GJ, Macarthur R, Delahay RJ, McDonald RA (2014) Density and abundance of badger social groups in England and Wales in 2011-2013. Scientific Reports: 4–3809.